DOCUMENT RESUME

ED 097 500 CE 002 268

AUTHOR	Ramsey-Klee, Diane M.; Richman, Vivian
TITLE	Continuation Studies of the Exploitation of the Narrative Sections of Navy Performance Evaluations for Senior Enlisted Personnel by Means of Content Analysis Techniques. Technical Report No. 3-74.
INSTITUTION	R-K Research and System Design, Malibu, Calif.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
PUB DATE	Jun 74
NOTE	129p.
EDRS PRICE	MF-$0.75 HC-$6.60 PLUS POSTAGE
DESCRIPTORS	Analysis of Variance; Classification; *Content Analysis; *Evaluation Methods; *Indexing; Information Retrieval; *Information Systems; Methods Research; *Military Personnel; Performance; Relevance (Information Retrieval); Reliability; Systems Analysis; Tables (Data)

ABSTRACT

The first task of the study was to develop valid, short-cut methods of indexing the narrative content of evaluation reports that would extract the differentiating information contained in evaluative comments simply, reliably, and with as good classification accuracy as the longer initial procedure. In the second task, the original inter-indexer reliability study was extended to clarify the issue of reliability of the complex, lengthy indexing procedure. Section 2 of the report reexamines the pilot study sample, the cross validation sample, and the generalization sample. Section 3 reviews the original content analysis methodology and includes a description of the two shortcut indexing methods that were devised. In Section 4 the performance of the two shortcut indexing methods in classifying the three experimental samples into correct criterion groups is compared with that of the original lengthy indexing procedure. Section 5 presents the results of the extension of the original inter-indexer reliability study. In Section 6 future areas of investigation are delineated. (Author/MW)

ED 097500

R-K RESEARCH AND SYSTEM DESIGN
3947 Ridgemont Drive
Malibu, California

Technical Report No. 3-74

CONTINUATION STUDIES OF THE EXPLOITATION OF THE NARRATIVE SECTIONS OF NAVY PERFORMANCE EVALUATIONS FOR SENIOR ENLISTED PERSONNEL BY MEANS OF CONTENT ANALYTIC TECHNIQUES

June 1974

Diane M. Ramsey-Klee, Ph.D.
Principal Investigator

Vivian Richman, M.L.S.
Research Associate

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

This research was sponsored by the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

Contract N00014-72-C-0231
Contract Authority Identification Number, NR 150-344

Reproduction in whole or in part is permitted
for any purpose of the United States Government.
Approved for public release; distribution unlimited.

CE 002 268

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report No. 3-74	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Continuation Studies of the Exploitation of the Narrative Sections of Navy Performance Evaluations for Senior Enlisted Personnel by Means of Content Analytic Techniques		5. TYPE OF REPORT & PERIOD COVERED Annual Technical Report (1 Mar 1973 - 31 Dec 1973) 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Diane M. Ramsey-Klee Vivian Richman		8. CONTRACT OR GRANT NUMBER(s) N00014-72-C-0231
9. PERFORMING ORGANIZATION NAME AND ADDRESS R-K Research and System Design 3947 Ridgemont Drive Malibu, California 90265		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR 042-04; RR 042-04-01 NR 150-344
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE June 1974 13. NUMBER OF PAGES 120
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Personnel technology; personnel selection; Navy Evaluation Reports; content analysis; reliability; stepwise discriminant analysis; performance evaluation; appraisal of performance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(U) The purpose of this continuing research effort is to develop content analytic techniques capable of extracting the differentiating information in narrative performance evaluations for senior enlisted personnel in order to aid selection boards in choosing the most qualified candidates for promotion. In the present study two tasks were performed. The first task was to try to develop valid, short-cut methods of indexing the narrative content of Evaluation Reports that would extract the differentiating information contained in evaluative comments in a simple but reliable fashion, hopefully achieving as

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

good or nearly as good classification accuracy in assigning individuals to three criterion groups by a stepwise discriminant analysis procedure as the longer, more complex indexing method developed earlier. In the second task an earlier inter-indexer reliability study was extended in order to elucidate more fully the issue of reliability of the complex, lengthy indexing procedure.

Two short-cut indexing methods were developed, one a rational condensation of the entire original hierarchy of 29 index terms into a new set of 15 compressed terms, and the other a 15-term subset of the original hierarchy of 29 terms chosen on the basis of their early selection by the stepwise discriminant analysis process. The two short-cut indexing methods, although not achieving the classification accuracy of the original lengthy indexing procedure which had more variables available for the stepwise discriminant analysis process, did, however, achieve an acceptable level of classification performance in comparison to the longer, more complex indexing methodology. Of the two short-cut methods, the rational condensation indexing method was preferred since it tracked the lengthy method more faithfully in the selection of discriminating variables. Further, the rational condensation method examines *all* of the information contained in a narrative performance evaluation whereas the statistically selected subset method ignores certain portions of the narrative text.

The key variables in differentiating between the performance of superlative chief petty officers and their slightly less qualified colleagues were the adjectives and adverbs that an evaluator uses to describe the performance of the individual that is being evaluated; the range of skills and abilities that an individual manifests; and the following specific demonstrated capabilities: Management and supervisory ability; skill in leading and directing his men; ability to organize his work area and to staff it properly; ability to plan his workload and take any corrective measures necessary to compensate for unforeseen obstacles to good performance; the ability to present an effective image of his work force to other components of the Navy and to the civilian community; skill in communicating effectively with others; a cooperative and responsive way of performing his job duties; a creative, resourceful, and innovative approach to his work; the drive and stamina to perform well under tiring or adverse circumstances; his level of intellectual functioning; professional and technical competence in his occupational specialty; his level of productivity and achievement; and recognition of his assets and potential by his subordinates, peers, and superior officers.

The results of the extension of the inter-indexer reliability study, using the lengthy complex indexing procedure, were very similar to the results of the original reliability study. In the original study product-moment correlation, kappa, and weighted kappa were the three statistics used to measure agreement among the four reliability indexers. Of the six possible pairwise comparisons between the four reliability indexers, the value of the various agreement statistics ranged from .64 to .88. In the extension of the reliability study, the various agreement statistics ranged from .48 to .83. However, it was felt that one of the data bases for the second reliability study contained a sample of narrative text more difficult to index than the first reliability study data base. But once again, the heartening finding was that in only six training sessions a quite respectable level of agreement among indexers was achieved. This is a significant finding because it suggests that Navy and civilian operational personnel also can be trained to consistently apply content analytic techniques.

CONTENTS

DD FORM 1473: REPORT DOCUMENTATION PAGE iii

LIST OF FIGURES vi

LIST OF TABLES vii

ACKNOWLEDGMENTS x

SUMMARY OF FINDINGS xi

SECTION 1 INTRODUCTION 1

SECTION 2 NATURE OF THE PILOT STUDY SAMPLE, THE CROSS VALIDATION SAMPLE, AND THE GENERALIZATION SAMPLE 5

SECTION 3 CONTENT ANALYSIS METHODOLOGY 11

Original Conceptual Approach 11
Original Indexing Vocabulary 11
Short-cut Indexing Methods 22

SECTION 4 COMPARISON OF THE SHORT-CUT INDEXING METHODS WITH THE ORIGINAL LENGTHY INDEXING PROCEDURE 29

AT Comparison 30
BT Comparison 52
CS Comparison 73
RM Comparison 77
Summary and Conclusions 84

SECTION 5 EXTENSION OF THE ORIGINAL INTER-INDEXER RELIABILITY STUDY 97

SECTION 6 FUTURE AREAS OF INVESTIGATION 111

REFERENCES 115

DISTRIBUTION LIST 117

LIST OF FIGURES

Figure 1	Evaluation Report Form NAVPERS 1616/8	6
Figure 2	Original Indexing Form Used in Performing the Content Analysis	12
Figure 3	Example of the Narrative Text for An Evaluation Section Showing the Original Indexing Decisions That Were Made. Factual Statements Requiring No Indexing Are Enclosed in Brackets. T = Total Number of Words in the Narrative Text.	16
Figure 4	The Original Indexing Form As It Was Filled Out to Record the Indexing Decisions Made in the Example of Narrative Text Shown in Figure 3	17
Figure 5	Indexing Form Used in Performing the Content Analysis for the Rational Condensation Short-cut Method	24
Figure 6	Indexing Form Used in Performing the Content Analysis for the Statistically Selected Subset Short-cut Method	27

LIST OF TABLES

Table 1	Range of Raw Marks, Range of T Scores, and Mean of T Scores on 19A-PERFORMANCE OF DUTY for the Three Criterion Groups in the Pilot Study Sample	7
Table 2	Distribution of the 225 Pilot Study Evaluation Reports Among the Three Criterion Groups for Each of the Two Occupational Specialties and for Both Occupations Combined	8
Table 3	Range of Raw Marks, Range of T Scores, and Mean of T Scores on 19A-PERFORMANCE OF DUTY for the Three Criterion Groups in the Cross Validation Sample	9
Table 4	Range of Raw Marks, Range of T Scores, and Mean of T Scores on 19A-PERFORMANCE OF DUTY for the Three Criterion Groups in the Generalization Sample	9
Table 5	Distribution of the 222 Cross Validation Sample Evaluation Reports Among the Three Criterion Groups for Each of the Two Occupational Specialties and for Both Occupations Combined	10
Table 6	Distribution of the 222 Generalization Sample Evaluation Reports Among the Three Criterion Groups for Each of the Two Occupational Specialties and for Both Occupations Combined	10
Table 7	Original Weighting Scale	14
Table 8	Definition of the 67 Quantitative Variables Derived from the Original Indexing Form	19
Table 9	Transformation of Original Weighting Scale	21
Table 10	Definition of the 23 Quantitative Variables Derived from the Rational Condensation Indexing Form	25
Table 11	Definition of the 21 Quantitative Variables Derived from the Statistically Selected Subset Indexing Form	28
Table 12	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Pilot Study AT's (N=144) - Evaluation Section*	31

Table 13 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Cross Validation AT's (N=138) - Evaluation Section* 35

Table 14 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Combined AT Samples (N=282) - Evaluation Section* 38

Table 15 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Pilot Study AT's (N=144) - Justification Section* 42

Table 16 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Cross Validation AT's (N=138) - Justification Section* 45

Table 17 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Combined AT Samples (N=282) - Justification Section* 49

Table 18 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Pilot Study BT's (N=80) - Evaluation Section* 53

Table 19 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Cross Validation BT's (N=84) - Evaluation Section* 56

Table 20 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Combined BT Samples (N=164) - Evaluation Section* 60

Table 21 Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Pilot Study BT's (N=80) - Justification Section* 63

Table 22	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Cross Validation BT's (N=84) - Justification Section*	67
Table 23	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Combined BT Samples (N=164) - Justification Section*	70
Table 24	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Generalization CS's (N=60) - Evaluation Section*	74
Table 25	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Generalization CS's (N=60) - Justification Section*	78
Table 26	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Generalization RM's (N=162) - Evaluation Section*	81
Table 27	Comparison of the Variables Selected at Each Step in the Stepwise Discriminant Analysis and the Classification Accuracy Achieved by the Three Indexing Procedures: *Generalization RM's (N=162) - Justification Section*	85
Table 28	Summary of the Best Classification Accuracy Achieved by the Three Indexing Procedures	89
Table 29	Results of the Kappa Analysis for the Six Pairwise Comparisons Between the Four Reliability Indexers and the Experienced Indexer in Selecting Index Terms for the Two Reliability Study Data Bases	101
Table 30	The Weighting Algorithm Used in Computing Weighted Kappa for Assessing the Level of Agreement in Assigning Numerical Weights to the Index Terms Selected in the Two Reliability Studies	104
Table 31	Results of the Correlational Analysis and the Weighted Kappa Analysis for the Six Pairwise Comparisons Between the Four Reliability Indexers and the Experienced Indexer in Assigning Numerical Weights to Each Index Term Selected for the Two Reliability Study Data Bases	107

ACKNOWLEDGMENTS

This ongoing research effort has continued to concern itself with developing content analytic techniques capable of extracting the differentiating information in narrative performance evaluations for senior enlisted personnel in order to aid selection boards in choosing the most qualified candidates for promotion. This technical report presents the findings resulting from the performance of two tasks, the first task an attempt to develop valid, short-cut methods of indexing the narrative content of Evaluation Reports, and the second task an extension of an earlier inter-indexer reliability study. These continuing investigations being conducted under the auspices of this project are sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research. The Navy Personnel Research and Development Center (NPRDC), San Diego, California, provided the data bases used in this research. The continuing support by ONR and the cooperation from NPRDC are gratefully acknowledged. Marshall J. Farr, Ph.D., and Joseph L. Young, Ph.D., Director and Assistant Director of Personnel and Training Research Programs, Office of Naval Research, have provided intellectual input and encouragement. David W. Robertson and Marjorie H. Royle, Personnel Utilization Research Department, Navy Personnel Research and Development Center, have given willingly and extensively of their time in obtaining the data bases and conferring about issues of experimental design and statistical methodology. Their guidance provided insight into potential applications for this research endeavor. To these four individuals I wish to express my special appreciation.

During the writing stage, this report was reviewed by David W. Robertson and Marjorie H. Royle. Their valuable suggestions have been incorporated into the final product.

The stepwise discriminant analyses were performed at the UCLA Health Sciences Computing Facility. This facility operates under the directorship of Wilfrid J. Dixon, Ph.D., and is sponsored by NIH Special Research Resources Grant RR-3. All of the agreement statistics calculated for the extension of the inter-indexer reliability study were performed on the Olivetti P602 microcomputer at R-K Research and System Design.

The conceptualization of the rational condensation short-cut indexing method was the work of Vivian Richman, M.L.S. Her extensive indexing experience and understanding of management theory contributed greatly to the development of this superior short-cut indexing technique. She also trained the four reliability indexers in the extension of the inter-indexer reliability study. Barry M. Richman, Ph.D., and Harold Koontz, Ph.D., offered constructive criticism and consultation in the area of appraisal of managers.

Mrs. Jonnie Handley assisted both the Principal Investigator and Mrs. Richman in the myriad tasks of coding for keypunching, quality control, data extraction, and statistical computation. She also served as one of the reliability indexers. Mrs. Carol J. Randall was responsible for the excellent typing of this technical report. She also served as one of the reliability indexers.

Diane M. Ramsey-Klee, Ph.D.
Principal Investigator

SUMMARY OF FINDINGS

The purpose of this continuing research investigation has been to complement the Navy Personnel Research and Development Center (NPRDC), San Diego, in their efforts to develop effective procedures for improving the validity of individual personnel selection decisions based on accurate measures of job-relevant performance. The workload facing selection boards is massive and to date narrative comments on the performance evaluation forms have not been exploited in any systematic manner because narrative text resists easy analysis. Yet there seems to be a great deal of differentiating information in these narrative comments that could substantially aid selection boards in choosing the most qualified candidates for promotion. Therefore, a series of studies has been conducted, all aimed at developing and refining content analytic techniques capable of extracting the differentiating information in narrative performance evaluations for senior enlisted personnel. In the study being reported here, two tasks were performed. The first task was to try to develop valid, short-cut methods of indexing the narrative content of Evaluation Reports that would extract the differentiating information contained in evaluative comments in a simple but reliable fashion, hopefully achieving as good or nearly as good classification accuracy in assigning individuals to three criterion groups by a stepwise discriminant analysis procedure as the longer, more complex indexing method developed earlier. In the second task an earlier inter-indexer reliability study was extended in order to elucidate more fully the issue of reliability of the complex, lengthy indexing procedure. Two short-cut indexing methods were developed, one a rational condensation of the entire original hierarchy of 29 index terms into a new set of 15 compressed terms, and the other a 15-term subset of the original hierarchy of 29 terms chosen on the basis of their early selection by the stepwise discriminant analysis process.

Comparison of the Short-cut Indexing Methods with the Original Lengthy Procedure

The two short-cut indexing procedures that were developed for this study compared favorably with the classification accuracy achieved by the original lengthy indexing procedure in the early steps of the stepwise discriminant analysis process, i.e., between Steps 1 and 10. Beyond Step 10 the lengthy indexing procedure, with its greater complement of available variables, typically displayed a superior classification performance as the stepwise discriminant analysis process continued to try to maximize its classification accuracy. In all of the comparisons that were made, the lengthy indexing procedure exceeded the better classification performance of the two short-cut index-dexing methods. However, since the lengthy indexing procedure provided more variables to the stepwise discriminant analysis process, it was expected that this method would demonstrate superior classification performance. There is other evidence that most of the discrimination which is achievable can be attributed to the variables selected early by the stepwise discriminant analysis process. In previous research with the lengthy indexing procedure, when the discriminant functions developed on one sample were used to classify a second cross validation sample, the classification performance of the lengthy procedure dropped markedly, typically from near perfect classification for the original sample to 65-70% classification accuracy for the cross validation sample. This level of cross validation classification accuracy was achieved early in the stepwise discriminant analysis process, typically by the fifth step. This important finding from a previous study indicated that the variables selected by the stepwise discriminant analysis program for the early steps in the discriminant analysis are crucial variables, playing a major role in differentiating among the three criterion groups.

Perhaps, then, a more meaningful comparison among the three indexing procedures is the classification performance that they achieved between Steps 10 and 20 in the stepwise discriminant analysis process, the range of steps at which the two short-cut methods attained their best classification accuracy. In the comparisons for four occupational specialties---AT's, BT's, CS's, and RM's, the classification performance of the three indexing procedures was similar between Steps 10 and 20, with the lengthy procedure typically having a slight but definite edge over the two short-cut methods. In some comparisons the rational condensation indexing method, at its best classification accuracy, demonstrated superior classification performance to the best performance of the statistically selected subset indexing method, but in other comparisons the statistically selected subset method performed better. In eight of the 16 comparisons that were made, the rational condensation method achieved better classification accuracy. In seven of the 16 comparisons, the statistically selected subset method attained better classification accuracy. In one comparison the two short-cut indexing methods performed equally well. Therefore, the criterion that was adopted to determine which of the two short-cut methods should be considered superior and elected as the preferred method for subsequent research studies was how well each short-cut method tracked the original lengthy indexing procedure in selecting variables into the discriminant function. Of the two short-cut indexing methods, the one that from the initial step more faithfully tracked the original lengthy indexing procedure in selecting variables into the discriminant function was the rational condensation method. Moreover, the rational condensation method examines *all* of the information contained in a narrative performance evaluation in contrast to the statistically selected subset method which takes into consideration only portions of the narrative text, thus subjecting it to more indexing error and inconsistency. Therefore, the rational condensation method was chosen as the preferred short-cut indexing method for further research investigations.

When the number of predictor variables is large in relation to the number of cases (the worst instance in this study being 67 variables for the lengthy indexing procedure as applied to the 60 generalization CS's), the solution achieved by the stepwise discriminant analysis algorithm, as in the case of multiple regression, may converge on a set of predictor variables that solves the classification problem perfectly for that particular sample, but may not constitute the same set of variables that might be selected for another sample or for another indexing procedure, a different set of variables also being able to achieve perfect or near perfect classification. Therefore, it is extremely interesting to note that for all three indexing procedures, the key variables selected early in the stepwise discriminant analysis process for the Evaluation Section were Total Number of 5 Weights (Excellent) and Total Number of 2 Weights (Poor). This was true for all comparisons made on the Evaluation Section of the Evaluation Report except for the 60 generalization CS's. This finding points up the need to cross validate the results of studies based on small N's where the number of predictor variables exceeds the number of cases in order to determine which discriminanting variables are constant over several samples. The conclusion that can be drawn from these findings is that the modifying adjectives and adverbs used by an evaluator to assess an individual's performance in the Evaluation Section of the Evaluation Report are key factors in distinguishing between superior performance and less stellar achievements, regardless of the occupational specialty being analyzed, with the exception of the 60-case generalization CS sample which constituted the worst case statistically for finding a valid, reproducible set of predictor variables.

When one examines the results for the Justification Section of the Evaluation Report, the findings are unequivocal. *Without exception* for all comparisons made, the first variable selected for the Justification Section was Total Number of Index Terms Used. This variable reflects the variety of specific areas of an individual's performance that the evaluator chose to comment on, and is measured by the number of different index terms chosen by the indexer to encompass the narrative content. This finding indicates that the range of skills and abilities that a chief petty officer manifests is a key factor in his superior performance as narrated by the evaluator in the Justification Section. Another finding, which corroborates the results of previous research, is that *without exception* better classification was achieved in the content analysis of the narrative comments in the Justification Section compared to the Evaluation Section, regardless of which of the three indexing procedures was employed.

The results from an earlier research study indicated that classification procedures based on the lengthy content analysis methodology should be tailored to specific occupations. The findings from the study being reported here substantiate the earlier research results and show that for each occupational specialty on a particular section of the Evaluation Report, the variables selected for at least two of the three indexing procedures were identical and not necessarily the same as those variables selected for a different occupational specialty. A summary enumeration of these key discriminating variables selected in the first 15 steps by the stepwise discriminant analysis procedure for each occupational specialty is given below.

AT's - Key Discriminating Variables for the Evaluation Section. The following list of variables was determined from the results of the combined AT analysis on the Evaluation Section (N=282). The key discriminating clusters of variables were Total Number of 5 Weights (Excellent); Total Number of 2 Weights (Poor); LEADERSHIP AND DIRECTING; TECHNICAL SKILLS/PROFESSIONAL AND TECHNICAL SKILLS; MANAGEMENT FUNCTIONS; Total Number of 3 Weights (Average); RESPONSIVENESS; COMMUNICATION; POTENTIAL; and DRIVE.

AT's - Key Discriminating Variables for the Justification Section. The following list of variables was determined from the results of the combined AT analysis on the Justification Section (N=282). The key discriminating clusters of variables were Total Number of Index Terms Used; sum of the weighted frequencies of the available set of variables for a particular indexing procedure; PRODUCTIVITY AND ACHIEVEMENT; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS/PROFESSIONALISM; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; STAFFING/ORGANIZATION AND STAFFING; INTELLECTUAL FUNCTIONING; ENDURANCE AND MOTIVATION/DRIVE; MANAGEMENT FUNCTIONS; REPUTE; Total Number of 2 Weights (Poor); COMMUNICATION; REPRESENTATION; and Total Number of Words in Text.

BT's - Key Discriminating Variables for the Evaluation Section. The following list of variables was determined from the results of the combined BT analysis on the Evaluation Section (N=164). The key discriminating clusters of variables were Total Number of 5 Weights (Excellent); Total Number of 2 Weights (Poor); Total Number of Index Terms Used; MANAGEMENT FUNCTIONS; COMMUNICATION; PROFESSIONALISM; RECOGNITION/REPUTE/ASSET TO THE NAVY/POTENTIAL; SKILLS AND ABILITIES; RESOURCEFULNESS/CREATIVITY AND INITIATIVE; REPRESENTATION; ORGANIZATION/STAFFING/ORGANIZATION AND STAFFING; and PRODUCTIVITY AND ACHIEVEMENT.

BT's - Key Discriminating Variables for the Justification Section. The following list of variables was determined from the results of the combined BT analysis on the Justification Section (N=164). The key discriminating clusters of variables were Total Number of Index Terms Used; LEADERSHIP AND DIRECTING; sum of the simple or weighted frequencies of the available set of variables for a particular indexing method; PRODUCTIVITY AND ACHIEVEMENT; RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; CREATIVITY AND INITIATIVE/INITIATIVE; AWARDS AND PUNISHMENT/ASSET TO THE NAVY/RECOGNITION; SKILLS AND ABILITIES; and Total Number of Words in Text.

CS's - Key Discriminating Variables for the Evaluation Section. The key discriminating clusters of variables for the 60 generalization CS's on the Evaluation Section were CONTROLLING/PLANNING-CONTROLLING; ASSET TO THE NAVY; MANAGEMENT FUNCTIONS; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of Index Terms Used; PLANNING; ENDURANCE/SERVICE MOTIVATION/DRIVE/ENDURANCE AND MOTIVATION; CREATIVITY AND INITIATIVE/INITIATIVE; ORGANIZATION AND STAFFING/ORGANIZATION; POTENTIAL; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS; LEADERSHIP AND DIRECTING; and REPRESENTATION.

CS's - Key Discriminating Variables for the Justification Section. The key discriminating clusters of variables for the 60 generalization CS's on the Justification Section were Total Number of Index Terms Used; PROFESSIONALISM/PROFESSIONAL AND TECHNICAL SKILLS; INITIATIVE/CREATIVITY AND INITIATIVE; COMMUNICATION; COOPERATION AND RESPONSIVENESS/COOPERATION; REPRESENTATION; STAFFING/ORFANIZATION AND STAFFING; POTENTIAL; Total Number of Words in Text; PRODUCTIVITY AND ACHIEVEMENT; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; and PLANNING.

RM's - Key Discriminating Variables for the Evaluation Section. The key discriminating clusters of variables for the 162 generalization RM's on the Evaluation Section were Total Number of 2 Weights (Poor); AWARDS AND PUNISHMENT/RECOGNITION/POTENTIAL/REPUTE/ASSET TO THE NAVY; MANAGEMENT FUNCTIONS; Total Number of 5 Weights (Excellent); COMMUNICATION; RELIABILITY AND DEPENDABILITY/CONDUCT AND ATTITUDE; INTELLECTUAL FUNCTIONING; Total Number of Index Terms Used; PRODUCTIVITY AND ACHIEVEMENT; REPRESENTATION; and RESPONSIVENESS.

RM's - Key Discriminating Variables for the Justification Section. The key discriminating clusters of variables for the 162 generalization RM's on the Justification Section were Total Number of Index Terms Used; sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of 4 Weights (Good); PRODUCTIVITY AND ACHIEVEMENT; DRIVE/ENDURANCE AND MOTIVATION/ENDURANCE; COOPERATION/RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; CONDUCT AND ATTITUDE/GROOMING AND ATTIRE; STAFFING/ORGANIZATION AND STAFFING; Total Number of 3 Weights (Average); REPUTE; PRO-

FESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS/PROFESSIONALISM; PLANNING; CREATIVITY AND INITIATIVE/INITIATIVE; and POTENTIAL.

There is very little difference between the original lengthy indexing procedure and the superior rational condensation short-cut indexing method in the time required to index and code Evaluation Reports containing brief narrative text. Only when the text becomes longer and requires more scrutiny and consideration by the indexer does the efficiency of the short-cut method become apparent. Over a large sample of Evaluation Reports, it is estimated that use of the rational condensation short-cut indexing method will save 25 to 50 percent of the indexing time required by the original lengthy indexing procedure. It is expected that the time required to count the number of words in the narrative text and to transfer this count and the indexing decisions to the indexing form and to generate the various quantitative variables will be approximately the same for both procedures. The time required for the rational condensation method to enter this information onto IBM coding forms preparatory to keypunching is estimated to be two-thirds of that required by the lengthy indexing procedure. Since only one punched card is needed to contain the variables extracted by the rational condensation content analysis compared to two punched cards for the original lengthy content analysis, the keypunching, verifying, and proofing time is cut in half. And since fewer card images have to be examined by the stepwise discriminant analysis procedure each time that a classification matrix is computed and printed, it is estimated that computer processing time is halved.

Three samples of Evaluation Reports, covering two contiguous years and representing four occupational specialties and three experimental content analysis procedures, have highlighted certain key variables as being crucial in differentiating between the performance of superlative chief petty officers and their slightly less qualified colleagues. These key variables are the adjectives and adverbs that an evaluator uses to describe the performance of the individual that is being evaluated; the range of skills and abilities that an individual manifests; and the following specific demonstrated capabilities: Management and supervisory ability; skill in leading and directing his men; ability to organize his work area and to staff it properly; ability to plan his workload and take any corrective measures necessary to compensate for unforeseen obstacles to good performance; the ability to present an effective image of his work force to other components of the Navy and to the civilian community; skill in communicating effectively with others; a cooperative and responsive way of performing his job duties; a creative, resourceful, and innovative approach to his work; the drive and stamina to perform well under tiring or adverse circumstances; his level of intellectual functioning; professional and technical competence in his occupational specialty; his level of productivity and achievement; and recognition of his assets and potential by his subordinates, peers, and superior officers.

Extension of the Original Inter-Indexer Reliability Study

The original plan for the extension of the inter-indexer reliability study was to select and train four more individuals in the complex indexing procedure and to have them independently index the same 48 Evaluation Reports that formed the indexing corpus for the first reliability study. However, the results from the first reliability study strongly suggested that additional training of the original reliability indexers aimed at clarifying the areas of confusion that were identified in the analysis of their indexing judgments most likely would raise their level of agreement. Consequently, both of these avenues of investigation were pursued. A revision of the original training manual was prepared by the experienced indexer and the principal investigator, an updated version that attempted to eliminate areas of confusion brought to light in analyzing the results of the first reliability study and which also included voluminous examples of how to handle difficult indexing decisions. This revision was used to train the four participants in the extension of the original reliability study.

Two new reliability indexers were engaged for the study, a male and a female, both in their sophomore year in college. The other two indexers participating in the study were inexperienced indexer A (with two years of college in the liberal arts) and inexperienced indexer B (with executive secretary experience) who also had participated in the first reliability study. All four of these individuals were trained intensively by the experienced indexer over the course of six training sessions using the updated version of the training manual and a corpus of training Evaluation Reports. The two new reliability indexers independently indexed the same 48 Evaluation Reports that were indexed in the first reliability study. These two individuals in essence were attempting to replicate the earlier results. Inexperienced indexer A and inexperienced indexer B were given a new and different set of 48 Evaluation Reports to index independently. This second corpus constitutes a randomized representative sample taken from the cross validation and generalization data bases. This second aspect of the reliability study was included in order to test the hypothesis that with additional training and indexing experience, the level of indexing agreement can be raised.

In summary, the conclusions that can be drawn from this extension of the original reliability study are that once again, in only six training sessions, a fairly respectable level of agreement was achieved on a very difficult content analysis task, the various agreement statistics that were computed ranging from .48 to .83. The two new reliability indexers (both college sophomores) who were attempting to replicate the results from the first study did not achieve as high a level of agreement with the experienced indexer as the three reliability indexers did in the initial study, probably because the two new indexers were less motivated and not as deeply involved in the second reliability study as inexperienced indexers A and B were in the first study conducted a year earlier. These latter two individuals are regular employees of R-K Research and System Design, performing a variety of clerical and technical assignments in addition to their role in the two reliability studies. Inexperienced indexer A in particular may have had additional unsuspected training in the content analysis methodology since one of her other assignments in this research was to enter the indexing decisions of the experienced indexer for the pilot study, cross validation, and generalization samples onto IBM coding forms for keypunching. Inexperienced indexer A's extended exposure to the logic of the indexing scheme in the context of preparing the coding forms may account for her superior performance in both reliability studies.

In that part of the second reliability study designed to test the hypothesis that with additional training and indexing experience the level of indexing agreement can be raised, the results were ambiguous. Neither inexperienced indexer A nor inexperienced indexer B was able to increase her level of agreement with the experienced indexer despite refresher training in the complex, lengthy indexing procedure and the challenge to try to outdo her previous performance. However, these two reliability indexers felt that the data base indexed by them in the second reliability study contained a sample of narrative text more difficult to index than the first reliability study data base, and this greater difficulty inherent in the narrative text may have masked any gain in indexing proficiency that might have been achieved by the additional training. Another possible explanation is that inexperienced indexers A and B may have already approached the upper boundary of their indexing skill, with additional training and experience contributing very little to increasing their level of agreement with the experienced indexer.

SECTION 1. INTRODUCTION

The purpose of this continuing research investigation has been to complement the Navy Personnel Research and Development Center (NPRDC), San Diego, in their efforts to develop effective procedures for improving the validity of individual personnel selection decisions based on accurate measures of job-relevant performance. NPRDC has an ongoing program to develop and exploit Navy enlisted performance evaluation formats which will be effective in holding down the pile-up of marks at the high end of the marking scale and in achieving a distribution of marks that tapers off sufficiently at the high end of the scale in order to permit greater differentiation, thus making evaluations more useful, especially when small selection opportunities are involved.[1,2,3] An accurate and timely measure of each individual's on-job performance is essential if valid decisions are to be made in selecting personnel for advancement, duty assignment, training, or quality retention. Such a measure is one of the best indications of how well the individual will perform in other or future assignments. However, effective use of performance measures is severely limited due to the lack of performance data in formats responsive to the needs of the decision makers. The problem is particularly acute when these decision makers are members of selection boards who must review in a short span of time narrative evaluation data for thousands of candidates. The seriousness of this problem can be attested to by the fact that some 14,000 candidates for promotion must be reviewed annually by the E8-E9 selection board. This number represents the top 75 percent of all candidates, the lower 25 percent having been eliminated by a screening procedure. The problem is even more serious at the lower pay grades. An E7 selection board was established in 1973 whose task is to review annually the records of some 20,000 enlisted candidates for promotion to chief petty officer. This number of candidates represents the top 50 percent of the eligible population, the bottom 50 percent having been eliminated by a screening procedure.

The workload facing these selection boards is massive and to date narrative comments on the performance evaluation forms have not been exploited in any systematic manner because narrative text resists easy analysis. Yet there seems to be a great deal of differentiating information in these narrative comments that could substantially aid selection boards in choosing the most qualified candidates for promotion. This, then, was the task that R-K Research and System Design took on in an initial pilot study.

In the pilot study of the narrative sections of Navy performance evaluations for senior enlisted personnel in Pay Grade E7, it was determined by content analytic techniques that it is possible to differentiate between the performance of typical and superlative chief petty officers based on the substantive content of Evaluation Reports.[4] The results of this pilot study strongly suggested that there are stable differences among the performance characteristics of chief petty officers in the various portions of the upper half of the marking scale on Performance of Duty that are reflected in narrative statements written by evaluators. Prior to embarking on the initial pilot study, it was assumed that differences in marks between the upper and *lower* halves of the marking scale were readily reflected in narrative statements. However, in order to address the study to the realities and difficulties facing selection boards, who must make their selections only from a uniformly high quality group of candidates, NPRDC provided R-K Research and System Design with a truncated data set comprising individuals marked only in the upper half of the marking scale. The sample data then were divided into three criterion groups---Upper, Middle, and Lower---corresponding to three segments of the upper half of the marking scale on Performance of Duty (the criterion variable). This truncated data set required a much more rigorous analytical approach than would have been required for a nontruncated data set.

The statistical analyses that were performed on the quantitative data extracted from the pilot study content analysis supported the hypothesis that narrative performance evaluations do contain information useful to personnel selection boards in discriminating between typical and superlative chief petty officers. The findings from the pilot study were considered to be provocative enough to warrant further investigation. Therefore, a second study was embarked upon to attempt to cross validate the pilot study results on new Evaluation Reports for senior enlisted men in the same two occupational specialties (AT's and BT's) that were represented in the pilot study sample and to extend the content analysis to Evaluation Reports for senior enlisted men in two different occupational specialties (CS's and RM's) than those investigated in the pilot study in order to test the generalizability of the content analytic techniques developed earlier.[5] As a further refinement, the cross validation and generalization samples of Evaluation Reports were analyzed without any knowledge of the individual's relative position in the upper half of the marking scale on Performance of Duty (the criterion variable). In the pilot study the criterion data were made available early in the study, thus introducing the possibility that this knowledge subconsciously might have influenced the content analysis that was performed. This factor was controlled for in the second study by withholding the criterion information until the content analysis of the narrative text had been completed.

In the second study a series of more sophisticated and comprehensive statistical analyses was performed on the quantitative data extracted from the content analysis, resulting in the following important findings. It was possible to index the cross validation sample in the blind, without knowledge of criterion group membership, and achieve as good classification accuracy as was achieved with the pilot study sample where criterion group membership was known to the indexer. Further, it was shown that better classification into the three criterion groups was achieved when the two occupational specialties represented in the pilot study sample and the cross validation sample were treated separately. These findings suggest that classification procedures based on the content analysis methodology developed in this research should be tailored to specific occupations. In addition, it was shown that the content analysis methodology developed initially on the pilot study sample consisting of AT's and BT's was generalizable to a new sample consisting of two different occupational specialties, viz., CS's and RM's.

Also of concern in the pilot study were the issues of reliability and trainability, although the scope of the small initial research effort did not permit these aspects to be studied in any substantial way. Therefore, in designing the second investigation these issues were dealt with by including a reliability study whose objectives were twofold: (1) to determine the level of agreement among four individuals all of whom independently would perform a content analysis of the same corpus of Evaluation Reports, and (2) to investigate if nonresearchers could be trained successfully to apply the complex content analysis methodology developed in the pilot study.[5] Product-moment correlation, kappa, and weighted kappa were the three statistics used to measure agreement among the four reliability indexers. Of the six possible pairwise comparisons between the four reliability indexers, the value of the various agreement statistics ranged from .64 to .88. The initial expectation in beginning this reliability study was that it would be extremely difficult to train nonresearch-oriented individuals to consistently index the narrative sections of Evaluation Report forms using the complex content analysis methodology that had been developed in the pilot study. The surprising result was that in only six training sessions a quite respectable level of agreement was achieved. This is a significant finding because it suggests that Navy and civilian operational personnel also can be trained to consistently apply content analytic techniques.

In the follow-on investigation to the pilot study and the second study being reported here, two tasks were performed. The first task was to try to develop valid, short-cut methods of indexing the narrative content of Evaluation Reports that would extract the differentiating information contained in evaluative comments in a simple but reliable fashion, hopefully achieving as good or nearly as good classification accuracy as the longer, more complex indexing procedure developed initially. In the second task the original inter-indexer reliability study was extended in order to elucidate more fully the issue of reliability of the complex, lengthy indexing procedure.

Section 2 of this report reiterates the nature of the pilot study sample, the cross validation sample, and the generalization sample. Section 3 reviews the original content analysis methodology that was used in both the pilot study and the second study; this section also includes a description of the two short-cut indexing methods that were devised. In Section 4 the performance of the two short-cut indexing methods in classifying the three experimental samples into correct criterion group is compared with that of the original lengthy indexing procedure. Section 5 presents the results of the extension of the original inter-indexer reliability study. In Section 6 future areas of investigation are delineated.

SECTION 2. NATURE OF THE PILOT STUDY SAMPLE, THE CROSS VALIDATION SAMPLE, AND THE GENERALIZATION SAMPLE

As a result of research conducted at the Navy Personnel Research and Development Center (NPRDC), San Diego, to develop experimental forms for evaluating personnel in Pay Grades E7 (Chief Petty Officer), E8 (Senior Chief Petty Officer), and E9 (Master Chief Petty Officer), a new evaluation report form---NAVPERS 1616/8---was introduced into operational use in January 1969 (see Figure 1). This form subsequently has been replaced by another form that can be scanned by an optical character reader; however, the content of the two forms is essentially the same.

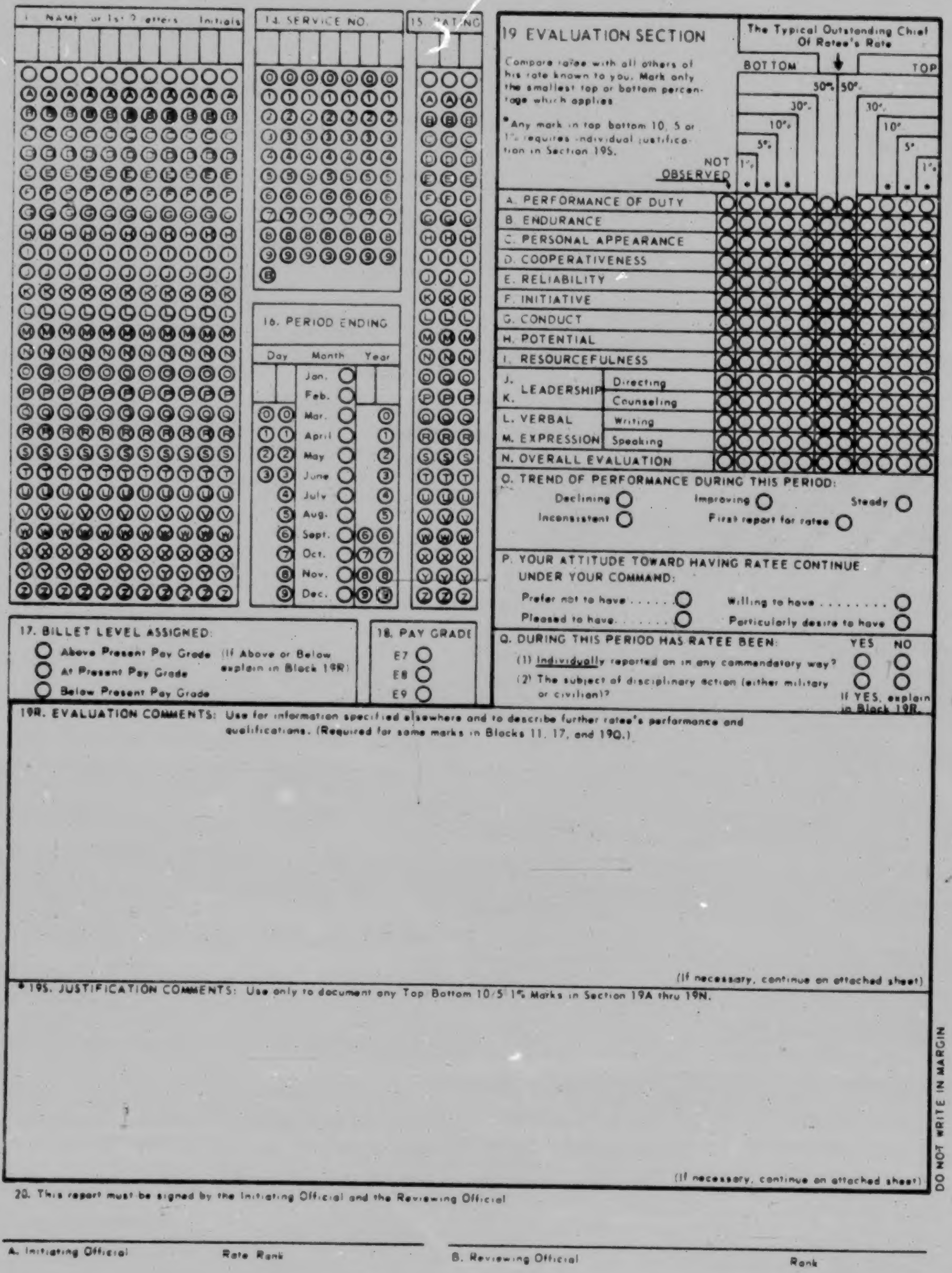
Section 19, Evaluation Section, of Evaluation Report Form NAVPERS 1616/8 is designed to permit the evaluator to *compare* the individual being evaluated with all others of his rate* known to the evaluator on 12 specific aspects of on-job performance. Evaluations are made by marking the column of the marking scale into which the evaluator decides that the individual falls for each of the 12 specific aspects of on-job performance plus an overall evaluation of the individual (for example, top 1% for superlative performance). Section 19R of this form provides space for the evaluator to write narrative evaluation comments to describe further the individual's performance and qualifications. Section 19S of this form provides space for the evaluator to write narrative justification comments and is required to support any marks assigned to the top or bottom 10, 5, or 1% columns of Section 19. Sections 19R and 19S are referred to as the narrative text of the Evaluation Report since they are the only portions of the report where the evaluator uses his own words to assess the on-job performance of the senior enlisted man whom he is evaluating.

In the pilot investigation, NPRDC selected a sample of 225 Evaluation Reports for senior enlisted personnel in Pay Grade E7 including 145 Aviation Electronics Technicians (AT's) and 80 Boiler Technicians (BT's). All 225 Evaluation Reports were drawn from the top half of the marking scale on 19A-PERFORMANCE OF DUTY located in the upper right quadrant of Evaluation Report Form NAVPERS 1616/8. The 19A-PERFORMANCE OF DUTY category was used in preference to 19N-OVERALL EVALUATION because standard scores (T Scores) were available only for 19A. The use of standard scores rather than raw marks permitted a more refined selection to be made of the three criterion groups used in the study. Since raw marks on 19A correlate very highly with raw marks on 19N, it was felt that little was sacrificed by not using the overall evaluation and that much was gained by using the purified T Scores on 19A. Only those Evaluation Reports from commands spreading their marks and submitting eight or more E7 and E8 reports were considered.

The pilot study sample of 225 Evaluation Reports was divided equally into three criterion groups---Upper, Middle, and Lower---corresponding to three continuous segments of the upper half of the marking scale on 19A-PERFORMANCE OF DUTY. Table 1 shows the range of raw marks on 19A for each of the three criterion groups in the pilot study sample as well as the range and mean of T Scores. These standardized scores have a mean of 50 and a standard deviation of 10. Standardization was accomplished by setting each unit command mean

---

* *Rate* is a Navy term which identifies an occupational specialty and pay grade.



1. NAME (1st [illegible] letters, Initials)

14. SERVICE NO.

15. RATING

16. PERIOD ENDING

Day Month Year

Jan. Feb. Mar. April May June July Aug. Sept. Oct. Nov. Dec.

19 EVALUATION SECTION

Compare ratee with all others of his rate known to you. Mark only the smallest top or bottom percentage which applies.

*Any mark in top bottom 10, 5 or 1% requires individual justification in Section 19S.

The Typical Outstanding Chief Of Ratee's Rate

BOTTOM 50% 30% 10% 5% 1%

TOP 50% 30% 10% 5% 1%

NOT OBSERVED

A. PERFORMANCE OF DUTY

B. ENDURANCE

C. PERSONAL APPEARANCE

D. COOPERATIVENESS

E. RELIABILITY

F. INITIATIVE

G. CONDUCT

H. POTENTIAL

I. RESOURCEFULNESS

J. LEADERSHIP Directing

K. LEADERSHIP Counseling

L. VERBAL EXPRESSION Writing

M. VERBAL EXPRESSION Speaking

N. OVERALL EVALUATION

O. TREND OF PERFORMANCE DURING THIS PERIOD:

Declining ○ Improving ○ Steady ○

Inconsistent ○ First report for ratee ○

P. YOUR ATTITUDE TOWARD HAVING RATEE CONTINUE UNDER YOUR COMMAND:

Prefer not to have . . . . . ○ Willing to have . . . . ○

Pleased to have. . . . . . ○ Particularly desire to have ○

Q. DURING THIS PERIOD HAS RATEE BEEN: YES NO

(1) Individually reported on in any commendatory way? ○ ○

(2) The subject of disciplinary action (either military or civilian)? ○ ○

If YES, explain in Block 19R.

17. BILLET LEVEL ASSIGNED:

○ Above Present Pay Grade

○ At Present Pay Grade

○ Below Present Pay Grade

(If Above or Below explain in Block 19R)

18. PAY GRADE

E7 ○

E8 ○

E9 ○

19R. EVALUATION COMMENTS: Use for information specified elsewhere and to describe further ratee's performance and qualifications. (Required for some marks in Blocks 11, 17, and 19Q.)

(If necessary, continue on attached sheet)

*19S. JUSTIFICATION COMMENTS: Use only to document any Top Bottom 10/5/1% Marks in Section 19A thru 19N.

(If necessary, continue on attached sheet)

DO NOT WRITE IN MARGIN

20. This report must be signed by the Initiating Official and the Reviewing Official

A. Initiating Official Rate Rank

B. Reviewing Official Rank

Figure 1. Evaluation Report Form NAVPERS 1616/8 (a 75 percent photo reduction of the original form)

TABLE 1

RANGE OF RAW MARKS, RANGE OF T SCORES, AND MEAN OF T SCORES ON 19A-PERFORMANCE OF DUTY FOR THE THREE CRITERION GROUPS IN THE PILOT STUDY SAMPLE

Criterion Group	Range of Raw Marks	Range of T Scores	Mean of T Scores
Upper	In the top 5% column or the top 1% column	59.3 to 74.2	64.68
Middle	In the top 10% column only	48.0 to 54.1	51.79
Lower	In the top 50% column or the top 30% column	33.8 to 39.7	38.85

equal to 50 and standardizing the total of E7 and E8 marks for each unit command. No cases from the bottom half of the marking scale on 19A were included in this study since there is no difficulty in differentiating these cases from the better performing personnel.

Table 2 shows the distribution of the 225 pilot study Evaluation Reports among the three criterion groups for each of the two occupational specialties and for both occupations combined. After the pilot study sample had been selected and analyzed, it was discovered that one Evaluation Report for an Aviation Antisubmarine Warfare Operator (AW) erroneously had been coded as an Aviation Electronics Technician (AT). This case was removed from the analysis by specific occupation but was left in the analysis for the total pilot study sample.

In the second study, NPRDC also selected the sample of Evaluation Reports to be analyzed. The same general procedures described above for selecting the pilot study sample were followed also in selecting the cross validation sample and the generalization sample, except that the forms were selected from a subsequent year's data pool. The cross validation sample consisted of 222 Evaluation Reports from the same two occupational specialties that were used in the pilot study, that is, Aviation Electronics Technician (AT) and Boiler Technician (BT). In addition, a generalization sample consisting of 222 Evaluation Reports was selected by NPRDC from two different specialties in order to ascertain the generalizability of the content analytic methodology developed in the pilot study. The two specialties from which the generalization sample was drawn were Commissaryman (CS) and Radioman (RM).

The cross validation sample of 222 Evaluation Reports and the generalization sample of 222 Evaluation Reports both were divided equally into the same three criterion groups---Upper, Middle, and Lower---taken from the upper half of the marking scale on 19A-PERFORMANCE OF DUTY as was the pilot study sample.

TABLE 2

DISTRIBUTION OF THE 225 PILOT STUDY EVALUATION REPORTS AMONG THE THREE CRITERION GROUPS FOR EACH OF THE TWO OCCUPATIONAL SPECIALTIES AND FOR BOTH OCCUPATIONS COMBINED

Occupational Specialty	Criterion Group			Total N
	Upper	Middle	Lower	
AT	49	39	56	144
AW*	0	1	0	1
BT	26	35	19	80
Total Sample	75	75	75	225

* This case erroneously was coded as an AT initially.

Table 3 shows the range of raw marks on 19A for each of the three criterion groups in the cross validation sample (AT's and BT's), the range of T Scores, and the mean of the T Scores for each criterion group. These same data for the generalization sample (CS's and RM's) are presented in Table 4. Tables 5 and 6 show the distribution of the 222 cross validation sample Evaluation Reports and the 222 generalization sample Evaluation Reports among the three criterion groups for each of the two occupational specialties represented in each sample and for both occupations combined. Actual criterion group membership for the cross validation sample and the generalization sample was known only to NPRDC until the content analysis of the narrative text had been completed. Consequently, the content analysis of these two samples was conducted in the blind without benefit of knowing to which criterion group each Evaluation Report belonged.

TABLE 3

RANGE OF RAW MARKS, RANGE OF T SCORES, AND MEAN OF T SCORES ON 19A-PERFORMANCE OF DUTY FOR THE THREE CRITERION GROUPS IN THE CROSS VALIDATION SAMPLE

Criterion Group	Range of Raw Marks	Range of T Scores	Mean of T Scores
Upper	In the top 5% column or the top 1% column	61.2 to 71.9	64.23
Middle	In the top 10% column only	48.2 to 55.9	52.54
Lower	In the top 50% column or the top 30% column	30.3 to 42.0	38.48

TABLE 4

RANGE OF RAW MARKS, RANGE OF T SCORES, AND MEAN OF T SCORES ON 19A-PERFORMANCE OF DUTY FOR THE THREE CRITERION GROUPS IN THE GENERALIZATION SAMPLE

Criterion Group	Range of Raw Marks	Range of T Scores	Mean of T Scores
Upper	In the top 5% column or the top 1% column	61.2 to 74.8	64.33
Middle	In the top 10% column only	48.2 to 56.2	52.50
Lower	In the top 50% column or the top 30% column	34.5 to 41.5	38.56

TABLE 5

DISTRIBUTION OF THE 222 CROSS VALIDATION SAMPLE EVALUATION REPORTS AMONG THE THREE CRITERION GROUPS FOR EACH OF THE TWO OCCUPATIONAL SPECIALTIES AND FOR BOTH OCCUPATIONS COMBINED

Occupational Specialty	Criterion Group			Total N
	Upper	Middle	Lower	
AT	45	44	49	138
BT	29	30	25	84
Total Sample	74	74	74	222

TABLE 6

DISTRIBUTION OF THE 222 GENERALIZATION SAMPLE EVALUATION REPORTS AMONG THE THREE CRITERION GROUPS FOR EACH OF THE TWO OCCUPATIONAL SPECIALTIES AND FOR BOTH OCCUPATIONS COMBINED

Occupational Specialty	Criterion Group			Total N
	Upper	Middle	Lower	
CS	19	16	25	60
RM	55	58	49	162
Total Sample	74	74	74	222

SECTION 3. CONTENT ANALYSIS METHODOLOGY

Original Conceptual Approach

In the pilot study, the narrative portions of the 75 Evaluation Reports for each of the three criterion groups were read in their entirety before formalizing the method of content analysis to be used. In this review the Evaluation Section and the Justification Section (19R and 19S) were considered separately. Borrowing from the field of information science, it seemed most appropriate to regard each narrative section as a short *document* that had been written by the evaluator in order to communicate to a selection board or to a detailer the potential that the individual being evaluated had for promotion and increased responsibility. Considered in this framework, the analysis task then becomes one of ascertaining what the *document* is about (content analysis), specification of the content by a set of descriptive labels (indexing), and organization of an indexing vocabulary (controlling the form and semantics of the descriptive labels by lexicon and/or rule).[6,7] In order for the content analysis to be valid, Fairthorne[8] cautions that two aspects must be taken into consideration: (a) what the *document* is about, and (b) the circumstances of the expected uses of the content analysis with respect to a particular task or problem. Fairthorne's advice was attended to in the design of the content analysis methodology in that the indexing vocabulary which was developed relates strongly to the ultimate use to which performance evaluations are put, that is, the selection for promotion of outstanding chief petty officers in the face of limited promotional opportunities.

Original Indexing Vocabulary

In reading the narrative portions of the 75 Evaluation Reports for each of the three criterion groups in the pilot study sample, it became apparent that the attributes and characteristics being evaluated for an individual related primarily to his potential as a manager and supervisor. Consequently, several references in the area of managerial behavior and practice[9,10,11] were consulted as an aid to the development of the indexing vocabulary used in this study. An initial vocabulary containing 41 descriptive labels was devised and used to test the adequacy and manageability of the indexing method on 20 Evaluation Reports not included in the pilot study sample but similar to them in content. As a result of this experience, the original set of 41 labels was condensed into a more generic set of 29 index terms. The original indexing form, incorporating the final vocabulary that was used in both the pilot study and the second study, is shown in Figure 2.

The top line of the original indexing form carries fields for an identifying number for each individual being evaluated, which criterion group he belongs to (used only in the pilot study since criterion data were withheld in the second study until the indexing had been completed), and whether the section being indexed is an Evaluation Section (19R) or a Justification Section (19S). The indexing form itself is divided into three major parts: MANAGEMENT FUNCTIONS, SKILLS AND ABILITIES, and PRODUCTIVITY AND ACHIEVEMENT. Under each of these headings there are more detailed terms, providing the indexer with a 3-level hierarchy of descriptive labels from which to choose.

ID No.__________ Criterion Group__________ Section__________

Index Term — Freq.

- MANAGEMENT FUNCTIONS
  - CONTROLLING
  - LEADERSHIP AND DIRECTING
  - ORGANIZATION
  - PLANNING
  - REPRESENTATION
  - STAFFING
  - USE OF COMMUNICATION
- SKILLS AND ABILITIES
  - COMMUNICATION
  - CONDUCT, INTEGRITY, AND PRIDE
  - COOPERATION
  - ENDURANCE
  - FLEXIBILITY
  - GROOMING AND ATTIRE
  - INITIATIVE
  - INTELLECTUAL FUNCTIONING
  - PROFESSIONALISM
  - RELIABILITY AND DEPENDABILITY
  - RESOURCEFULNESS
  - RESPONSIVENESS
  - TECHNICAL SKILLS
- PRODUCTIVITY AND ACHIEVEMENT
  - AWARDS AND PUNISHMENT
  - DRIVE
    - SERVICE MOTIVATION
  - POTENTIAL
  - REPUTE
    - ASSET TO THE NAVY

FREQUENCY COUNTS: 5 ______, 4 ______, 3 ______, 2 ______, 1 ______

TOTAL NUMBER OF WORDS ______ TOTAL NUMBER OF INDEX TERMS ______

Figure 2. Original Indexing Form Used in Performing the Content Analysis

The first section of the original indexing form includes seven specific MANAGEMENT FUNCTIONS that many authorities on management practice agree are the characteristic *duties* of all managers.[9,10,11] Although some authorities believe that there are more, less, or different functions performed by managers, these seven functions were selected because they are representative of the duties that chief petty officers actually perform.

The second section of the original indexing form contains index terms for 13 specific SKILLS AND ABILITIES considered to be important by Navy supervisory personnel in performing effectively as a chief petty officer. While some authorities on management practice consider making a judgment about whether or not an individual possesses a skill, quality, or ability to be a subjective process, Navy evaluators do repeatedly call out these specific qualities in their narrative evaluations because many of these qualities are dimensions on which they mark the individual in Section 19 of the Evaluation Report. The first section of the original indexing form---MANAGEMENT FUNCTIONS---deals with how an individual performs his managerial functions and is result oriented, while the second section---SKILLS AND ABILITIES---contains index terms that relate to an individual's characteristics and qualities which, if used, may help him achieve good results.

The third section of the original indexing form---PRODUCTIVITY AND ACHIEVEMENT---is the most result-oriented section of the indexing hierarchy. Here are included the measures of overall performance. DRIVE and SERVICE MOTIVATION (a specific type of drive) are included in this section since drive is considered to be one of the more important variables leading to success. POTENTIAL also is included here since potential is a measure of future performance. AWARDS AND PUNISHMENT, REPUTE, and ASSET TO THE NAVY represent acknowledgments of an individual's performance, either positive or negative acknowledgment.

Each sentence of narrative text in the pilot study sample and the cross validation and generalization samples was read carefully and, where appropriate, divided into segments corresponding to the assignment of specific index terms. However, it is not enough to simply label a narrative statement with the most appropriate index term since the statement may have been a highly positive, quite positive, neutral, quite negative, or highly negative one. For example, in order to differentiate between the individual who plans superbly and the individual who plans inadequately, a weighting scale was devised to be applied to each index term that is used (see Table 7). The original weighting scale contains five numerical values ranging from 5 (the positive end of the scale) to 1 (the negative end of the scale). Under each numerical value in Table 7 there are listed samples of adjectives or adverbs that may be used by the evaluator to describe an individual's performance. These lists of words provide clues to the indexer as to which numerical value to assign to an index term. As a simple example, if the evaluator commented that the individual was highly cooperative, this statement would be indexed as COOPERATION and assigned a weight of 4 since *highly* is listed as an example under numeral 4 in Table 7.

The narrative text of each Evaluation Report was read, segmented into distinct statements, and each statement was then assigned one or more index terms from the set of 29 possible choices shown in Figure 2. Each term selected was also assigned a numerical weight from 1 to 5 depending upon the nature

TABLE 7
ORIGINAL WEIGHTING SCALE

5 excellent	4 good	3 average	2 poor	1 poorest
superlative	comparative		comparative	superlative
best	better than most	average	not as good as most	worst
		EXAMPLES		
above reproach beyond reproach boundless exceptional extraordinary extremely finest flawless greatest highest ideal little to be desired limitless maximum most never outstanding paramount perfect profound sterling superb superior surpassed by none top/topnotch unimpeachable unique unlimited unmatched utmost without equal without exception 100%	above average better commendable complete deep definitely easily effective efficient eloquent eminent exceeds excels exemplary expeditious experienced expertise extensive favorable great high/highly immaculate immensely impeccable impressive innate inspires instills invaluable keen laudable leading marked meticulously model much noteworthy particularly rapidly 4-good (Cont.) rare remarkable significantly skillful smoothly solid strongly surpassed thorough tremendous truly unstinting valuable vast very	adequate aptly capable competent generally moderate satisfactory sufficiently usually	declining quality deficiency detrimental fair in need of insufficient lack of lower than average lowering of negatively spotty unfortunate unwisely weak in with the exception of	bottom least lowest

NOTE: AWARDS AND PUNISHMENT is assigned a weight of either 5 or 1.

of the adjectives or adverbs used as modifiers in the statement. The following examples will make more explicit the indexing procedure that was followed originally.

Example 1. "BTC has an excellent working and practical knowledge of the PMS System/but has a tendency to be lax in the administrative phase of the system."

This sentence was segmented into two parts. The first part was indexed as TECHNICAL SKILLS and assigned a weight of 5. The second part was indexed as MANAGEMENT FUNCTIONS and assigned a weight of 2.

Example 2. "Chief XX was relieved of his duties as the ship's Oil King after serving in this capacity for approximately two months./ He was removed from this billet because of his lack of professional knowledge/and technical know-how in the art of refueling."

This portion of narrative text was divided into three segments for indexing purposes. Segment 1 was indexed as AWARDS AND PUNISHMENT and assigned a weight of 1. Segment 2 was indexed as PROFESSIONALISM and assigned a weight of 2. Segment 3 was indexed as TECHNICAL SKILLS and assigned a weight of 2.

Example 3. "He is able to direct the efforts of Line Personnel in an efficient and effective manner;/this is reflected in CPO XX by a multiple of exceptional qualities."

This sentence was segmented into two parts. The first part was indexed as LEADERSHIP AND DIRECTING and assigned a weight of 4. The second part was indexed as SKILLS AND ABILITIES and assigned a weight of 5.

Example 4. "His natural abilities/and responsible approach to recruiting/ have enabled him to outperform his contemporaries."

This sentence was segmented into three parts. The first part was indexed as SKILLS AND ABILITIES and assigned a weight of 3. The second part was indexed as RELIABILITY AND DEPENDABILITY and assigned a weight of 3. The third part was indexed as PRODUCTIVITY AND ACHIEVEMENT and assigned a weight of 4.

Figure 3 shows an example of the complete narrative text written in an Evaluation Section as indexed originally. The index terms that were selected by the indexer have been recorded above each segment of text and the indexing weights that were assigned appear directly after each term. Factual statements requiring no indexing were enclosed in brackets. The number of words in the narrative text were counted and recorded at the bottom of the text by the indexer.

After all of the narrative text for either an Evaluation Section or a Justification Section of an Evaluation Report was indexed, the weights corresponding to each term were written onto the original indexing form to the right of the appropriate index term (see Figure 4). Thus there may have been two instances of mention of the individual's INTELLECTUAL FUNCTIONING, the first mention given a weight of 3 and the second a weight of 4. To the right of INTELLECTUAL FUNCTIONING on the indexing form for this person would be written the following string of weights: 3,4. Then to the far right on the indexing form

INT FUN 3 MAN FUN 4 PROD + ACH 5
[redacted] is an intelligent and proficient Petty Officer, who performs his duties
PLAN 3 20 ORG 3 ORG 3 L+D 3
in an outstanding manner. His ability to plan, organize, coordinate and super-
PROD + ACH 3
vise have been ably demonstrated by his performance as Recruit Company Command-
REL + DEP 3 REL + DEP 3 40 INT FUN 4
er. He is dependable, trustworthy, and exhibits mature judgment in disposing
CA + AT 3
of problems which occur within his company. [redacted]'s military appearance and
60 COND 4 COND 3
neatness of person and dress denote great pride. He is cheerful, highly mo-
DRIV 4 COOP 5
tivated, and gets along exceptionally well with others. His command of the
COMM 4 80
English language, both orally and written is above average. [redacted] is highly
POT 4 100
recommended for E-8. [He has been in Water Survival and Hygiene Division
POT 4
only for a short period of time.] He has shown a great potential towards being
120 DRIV 3
a swimming instructor and is practicing on his own time to qualify for Senior
Life Saver.

T = 132

Figure 3. Example of the Narrative Text for An Evaluation Section Showing the Original Indexing Decisions That Were Made. Factual Statements Requiring No Indexing Are Enclosed in Brackets. T = Total Number of Words in the Narrative Text.

ID No. 1000 Criterion Group Unknown at time of indexing Section Eval. (19 R)

Index Term		Freq.
MANAGEMENT FUNCTIONS	4	1
CONTROLLING		
LEADERSHIP AND DIRECTING	3	1
ORGANIZATION	3, 3	2
PLANNING	3	1
REPRESENTATION		
STAFFING		
USE OF COMMUNICATION		
SKILLS AND ABILITIES		
COMMUNICATION	4	1
CONDUCT, INTEGRITY, AND PRIDE	4, 3	2
COOPERATION	5	1
ENDURANCE		
FLEXIBILITY		
GROOMING AND ATTIRE	3	1
INITIATIVE		
INTELLECTUAL FUNCTIONING	3, 4	2
PROFESSIONALISM		
RELIABILITY AND DEPENDABILITY	3, 3	2
RESOURCEFULNESS		
RESPONSIVENESS		
TECHNICAL SKILLS		
PRODUCTIVITY AND ACHIEVEMENT	5, 3	2
AWARDS AND PUNISHMENT		
DRIVE	4, 3	2
SERVICE MOTIVATION		
POTENTIAL	4, 4	2
REPUTE		
ASSET TO THE NAVY		

FREQUENCY COUNTS: 5 2, 4 7, 3 11, 2 ____, 1 ____

TOTAL NUMBER OF WORDS 132 TOTAL NUMBER OF INDEX TERMS 13

Figure 4. The Original Indexing Form As It Was Filled Out to Record the Indexing Decisions Made in the Example of Narrative Text Shown in Figure 3

under the column headed "Freq." would be written "2", indicating that this index term had been used two times in indexing that particular section of narrative text.

At the bottom of the original indexing form there is a line labeled FREQUENCY COUNTS. After all of the weights assigned to the index terms selected for a section of narrative text (19R or 19S) had been entered on the indexing form, all of the 5 weights were counted and the sum was entered to the right of 5 on the FREQUENCY COUNTS line. The same procedure was followed for entering the frequency count of 4 weights, 3 weights, 2 weights, and 1 weights. The final step in completing the original indexing form was to transfer the total number of words written at the bottom of the narrative text and to count the total number of index terms selected from the set of 29 possibilities.

In order to increase the likelihood of consistent usage of the indexing vocabulary, a definition was written for each of the 29 index terms. Koontz and O'Donnell's *Principles of Management*[9] was relied upon heavily in defining the management-oriented terms listed in Figure 2. Also contributing to the formulation of the definitions for the 29 index terms was the way that Navy evaluators actually referred to these concepts in narrative text. These definitions were consulted frequently during the indexing process. Indexing of the pilot study sample and the cross validation and generalization samples was performed by one experienced indexer who also had conceptualized the content of the indexing vocabulary and had prepared the definitions of the 29 terms. As part of a concomitant study to ascertain the reliability of this content analysis methodology, a training manual was developed for use by the four reliability indexers who participated in the original reliability study. This training manual was included in its entirety in an earlier technical report.[5] The manual incorporates an alphabetical dictionary of the 29 index terms. The dictionary definition for each term is followed by extensive examples of correct indexing usage of the term and the proper assignment of weights.

A set of 67 quantitative variables was derived from the original indexing form used in the content analysis (see Table 8). The first 29 variables reflect the simple frequency with which each index term was used to index a particular section of narrative text. Variable 30 is the sum of these 29 frequencies. Variables 31 through 59 represent the *weighted* frequency of each index term used to index a particular section of narrative text. For example, suppose that the index term CONTROLLING was used twice. The first time that it was used it was assigned a weight of 4; the second time that it was used it was assigned a weight of 3. The weighted frequency then for CONTROLLING would be $4\times1 + 3\times1 = 7$. The simple frequency for this same example would be $1 + 1 = 2$. Variable 60 is similar to Variable 30 in that it is the sum of the 29 weighted frequencies.

Variables 61 through 65 represent the frequency counts over the entire original indexing form for all 5 weights, 4 weights, 3 weights, 2 weights, and 1 weights. Variable 66 is the total number of words in the section of narrative text that was indexed. Variable 67 is the total number of index terms of the 29 available that were used to index the section of narrative text.

Profiles or vectors of these 67 values then were prepared for all of the Evaluation Reports contained in each sample. Separate profiles were compiled for the Evaluation and Justification Sections of each Evaluation Report. If

TABLE 8

DEFINITION OF THE 67 QUANTITATIVE VARIABLES DERIVED FROM THE ORIGINAL INDEXING FORM

Number of Variable	Description of Variable
1	Frequency of Mention of MANAGEMENT FUNCTIONS
2	Frequency of Mention of CONTROLLING
3	Frequency of Mention of LEADERSHIP AND DIRECTING
4	Frequency of Mention of ORGANIZATION
5	Frequency of Mention of PLANNING
6	Frequency of Mention of REPRESENTATION
7	Frequency of Mention of STAFFING
8	Frequency of Mention of USE OF COMMUNICATION
9	Frequency of Mention of SKILLS AND ABILITIES
10	Frequency of Mention of COMMUNICATION
11	Frequency of Mention of CONDUCT, INTEGRITY, AND PRIDE
12	Frequency of Mention of COOPERATION
13	Frequency of Mention of ENDURANCE
14	Frequency of Mention of FLEXIBILITY
15	Frequency of Mention of GROOMING AND ATTIRE
16	Frequency of Mention of INITIATIVE
17	Frequency of Mention of INTELLECTUAL FUNCTIONING
18	Frequency of Mention of PROFESSIONALISM
19	Frequency of Mention of RELIABILITY AND DEPENDABILITY
20	Frequency of Mention of RESOURCEFULNESS
21	Frequency of Mention of RESPONSIVENESS
22	Frequency of Mention of TECHNICAL SKILLS
23	Frequency of Mention of PRODUCTIVITY AND ACHIEVEMENT
24	Frequency of Mention of AWARDS AND PUNISHMENT
25	Frequency of Mention of DRIVE
26	Frequency of Mention of SERVICE MOTIVATION
27	Frequency of Mention of POTENTIAL
28	Frequency of Mention of REPUTE
29	Frequency of Mention of ASSET TO THE NAVY
30	Sum of Variables 1 through 29
31	Weighted Frequency of Mention of MANAGEMENT FUNCTIONS
32	Weighted Frequency of Mention of CONTROLLING
33	Weighted Frequency of Mention of LEADERSHIP AND DIRECTING
34	Weighted Frequency of Mention of ORGANIZATION
35	Weighted Frequency of Mention of PLANNING
36	Weighted Frequency of Mention of REPRESENTATION

(Continued)

TABLE 8 (CONT.)

DEFINITION OF THE 67 QUANTITATIVE VARIABLES DERIVED FROM THE ORIGINAL INDEXING FORM

Number of Variable	Description of Variable
37	Weighted Frequency of Mention of STAFFING
38	Weighted Frequency of Mention of USE OF COMMUNICATION
39	Weighted Frequency of Mention of SKILLS AND ABILITIES
40	Weighted Frequency of Mention of COMMUNICATION
41	Weighted Frequency of Mention of CONDUCT, INTEGRITY, AND PRIDE
42	Weighted Frequency of Mention of COOPERATION
43	Weighted Frequency of Mention of ENDURANCE
44	Weighted Frequency of Mention of FLEXIBILITY
45	Weighted Frequency of Mention of GROOMING AND ATTIRE
46	Weighted Frequency of Mention of INITIATIVE
47	Weighted Frequency of Mention of INTELLECTUAL FUNCTIONING
48	Weighted Frequency of Mention of PROFESSIONALISM
49	Weighted Frequency of Mention of RELIABILITY AND DEPENDABILITY
50	Weighted Frequency of Mention of RESOURCEFULNESS
51	Weighted Frequency of Mention of RESPONSIVENESS
52	Weighted Frequency of Mention of TECHNICAL SKILLS
53	Weighted Frequency of Mention of PRODUCTIVITY AND ACHIEVEMENT
54	Weighted Frequency of Mention of AWARDS AND PUNISHMENT
55	Weighted Frequency of Mention of DRIVE
56	Weighted Frequency of Mention of SERVICE MOTIVATION
57	Weighted Frequency of Mention of POTENTIAL
58	Weighted Frequency of Mention of REPUTE
59	Weighted Frequency of Mention of ASSET TO THE NAVY
60	Sum of Variables 31 through 59
61	Total Number of 5 Weights
62	Total Number of 4 Weights
63	Total Number of 3 Weights
64	Total Number of 2 Weights
65	Total Number of 1 Weights
66	Total Number of Words in Narrative Text
67	Total Number of Index Terms Used

certain index terms were not used at all in indexing the Evaluation Section narrative or the Justification Section narrative, they were given a value of zero in the profile. This practice raised an important theoretical issue. Is it more damaging not to say anything about an individual's performance in a particular area than to damn him with qualified praise? A statement such as the following was assigned a weight of 2: "With more time and conscientious effort, he should realize a greater potential." This evaluation of the individual's potential seems more negative than not to have commented at all about his potential.

As a result of these considerations, the weighting scale that had been used in the original indexing of Variables 31 through 59 was transformed in order to place "no comment" between positive comments and negative comments. Table 9 shows the conversion that was used. A constant of 10 was added to the weighted frequency of Variables 31 through 59 in order to avoid the incidence of any negative input values in the subsequent statistical computations.

All profiles were transformed to the new weighting scale and entered onto IBM coding forms in preparation for keypunching. The criterion data and occupational specialty codes were known for the pilot study sample and were included on the coding forms. However, all of the coding forms for the cross validation sample and the generalization sample were sent to the Navy Personnel Research and Development Center in San Diego where the criterion data and occupational specialty codes were added to the coding forms and then returned to R-K Research and System Design for keypunching at UCLA. Card decks for each of the three samples were assembled in six parts: (1) Upper Criterion Group - Evaluation Section, (2) Middle Criterion Group - Evaluation Section, (3) Lower Criterion Group - Evaluation Section, (4) Upper Criterion Group - Justification Section, (5) Middle Criterion Group - Justification Section, and (6) Lower Criterion Group - Justification Section.

TABLE 9

TRANSFORMATION OF ORIGINAL WEIGHTING SCALE

Original Weights	Transformed Weights
5 (Excellent)	3 (Excellent)
4 (Good)	2 (Good)
3 (Average)	1 (Average)
	0 (No Comment)
2 (Poor)	-1 (Poor)
1 (Poorest)	-2 (Poorest)
0 (No Comment)	

Short-cut Indexing Methods

Two approaches to streamlining the original lengthy indexing procedure were devised. In the first approach the original hierarchy of 29 index terms was compressed into a rational condensation consisting of 15 terms. The rationale for this condensation grew out of extensive indexing experience and is based on management theory. The compression was achieved by combining those terms in the original hierarchy that logically belong together in management practice[9] or that tended to be confused with each other in the actual indexing of the pilot study, cross validation, and generalization data bases. This approach, called the *rational condensation*, includes all of the information contained in the original set of 29 index terms, but extracts this information in a more efficient, less confusing, and simpler fashion.

The second approach to streamlining the complex indexing methodology, called *statistically selected subset*, capitalized on the findings resulting from the various stepwise discriminant analyses that were performed originally on the pilot study, cross validation, and generalization samples. Plots of the classification accuracy achieved over the history of the discriminant analysis procedure revealed that the most useful information in discriminating between superior chief petty officers and their slightly less qualified colleagues is contained in the variables selected initially. Therefore, a subset of approximately one-third of the initial set of 67 quantitative variables derived from the original indexing form was determined, based on the order in which these variables were selected into the discriminant functions for the four occupational specialties represented in the pilot study, cross validation, and generalization data bases, i.e., AT's, BT's, CS's, and RM's.

Rational Condensation Method. In the rational condensation method, the original hierarchy of 29 index terms was compressed into 15 terms in the following fashion. The seven specific index terms comprising the MANAGEMENT FUNCTIONS section of the original hierarchy (see Figure 2) were collapsed into a new set of four index terms. MANAGEMENT FUNCTIONS remained as the major heading for this section, but CONTROLLING was combined with PLANNING and renamed PLANNING-CONTROLLING. LEADERSHIP AND DIRECTING remained unchanged as did REPRESENTATION. ORGANIZATION was combined with STAFFING and called ORGANIZATION AND STAFFING. USE OF COMMUNICATION was subsumed under COMMUNICATION in the SKILLS AND ABILITIES section of the hierarchy.

The second section of the hierarchy, that dealing with specific SKILLS AND ABILITIES, was condensed from 13 original index terms to seven revised terms. The term COMMUNICATION was expanded to include USE OF COMMUNICATION. A new term called CONDUCT AND ATTITUDE was created to encompass the following original index terms---CONDUCT, INTEGRITY, AND PRIDE; FLEXIBILITY; GROOMING AND ATTIRE; and RELIABILITY AND DEPENDABILITY. COOPERATION was combined with RESPONSIVENESS and renamed COOPERATION AND RESPONSIVENESS. The old terms INITIATIVE and RESOURCEFULNESS were combined to create a new term called CREATIVITY AND INITIATIVE. Another new term entitled ENDURANCE AND MOTIVATION subsumed the old term ENDURANCE from the SKILLS AND ABILITIES section of the original hierarchy and the old terms DRIVE and SERVICE MOTIVATION from the PRODUCTIVITY AND ACHIEVEMENT section of the original hierarchy. INTELLECTUAL FUNCTIONING remained unchanged. PROFESSIONALISM was combined with TECHNICAL SKILLS and renamed PROFESSIONAL AND TECHNICAL SKILLS.

The PRODUCTIVITY AND ACHIEVEMENT section of the compressed hierarchy now contains only one specific term---RECOGNITION---a consolidation of the old terms AWARDS AND PUNISHMENT, POTENTIAL, REPUTE, and ASSET TO THE NAVY, the old terms DRIVE and SERVICE MOTIVATION having been moved to the SKILLS AND ABILITIES section of the hierarchy and subsumed under the new term ENDURANCE AND MOTIVATION. Thus, the new hierarchy contains all of the concepts represented in the original indexing hierarchy but uses them in a more condensed and less confusing fashion.

Figure 5 portrays the indexing form that was developed for use with the rational condensation method. From this list of 15 compressed index terms, 23 quantitative variables were derived (see Table 10). The first 15 variables represent the *weighted* frequency of each index term used to index a particular section of narrative text using the original weighting scale shown in Table 7. The simple frequency of each index term was not employed because the weighted frequency provides more information. Variable 16 is the sum of the 15 weighted frequencies.

Variables 17 through 21 represent the frequency counts over the entire rational condensation indexing form for all 5 weights, 4 weights, 3 weights, 2 weights, and 1 weights. Variable 22 is the total number of index terms of the 15 available that were used to index the section of narrative text. Variable 23 is the total number of words in the section of narrative text that was indexed.

Statistically Selected Subset Method. In the statistically selected subset method, a subset of 15 terms was selected from the 29 original index terms, based on the order in which the original set of terms was selected by the stepwise discriminant analysis procedure. The 15 terms comprising this reduced, short-cut indexing procedure are those terms selected early by the discriminant analysis procedure for all four occupational specialties represented in the pilot study, cross validation, and generalization samples.

In the MANAGEMENT FUNCTIONS section of the original hierarchy (see Figure 2), the main heading was selected as well as four specific index terms under the MANAGEMENT FUNCTIONS heading---LEADERSHIP AND DIRECTING, PLANNING, REPRESENTATION, and STAFFING. In the SKILLS AND ABILITIES section of the original hierarchy, again the main heading was retained. Under this heading four specific skills and abilities were selected---COMMUNICATION, PROFESSIONALISM, RESPONSIVENESS, and TECHNICAL SKILLS. In the PRODUCTIVITY AND ACHIEVEMENT section of the original hierarchy, the main heading also was retained and under this heading the following specific terms were selected---DRIVE, POTENTIAL, REPUTE, and ASSET TO THE NAVY. The 14 original index terms that were *not* included in the statistically selected subset are the following: CONTROLLING; ORGANIZATION; USE OF COMMUNICATION; CONDUCT, INTEGRITY, AND PRIDE; COOPERATION; ENDURANCE; FLEXIBILITY; GROOMING AND ATTIRE; INITIATIVE; INTELLECTUAL FUNCTIONING; RELIABILITY AND DEPENDABILITY; RESOURCEFULNESS; AWARDS AND PUNISHMENT; and SERVICE MOTIVATION. These 14 terms were the least differentiating in that they were selected by the stepwise discriminant analysis procedure very late in the process and added very little to the classification accuracy of the discriminant function. Of these 14 terms the seven that were the most inferior in their discriminatory power were, in order from worst to better, GROOMING AND ATTIRE, RELIABILITY AND DEPENDABILITY, USE OF COMMUNICATION, CONTROLLING, ORGANIZATION, ENDURANCE, and INITIATIVE.

ID No.__________ Criterion Group_______________ Section____________

Index Term		Weighted Frequency
MANAGEMENT FUNCTIONS		
LEADERSHIP AND DIRECTING		
ORGANIZATION AND STAFFING		
PLANNING-CONTROLLING		
REPRESENTATION		
SKILLS AND ABILITIES		
COMMUNICATION		
CONDUCT AND ATTITUDE		
COOPERATION AND RESPONSIVENESS		
CREATIVITY AND INITIATIVE		
ENDURANCE AND MOTIVATION		
INTELLECTUAL FUNCTIONING		
PROFESSIONAL AND TECHNICAL SKILLS		
PRODUCTIVITY AND ACHIEVEMENT		
RECOGNITION		
	SUM OF WEIGHTED FREQUENCIES	

FREQUENCY COUNTS: 5 ______, 4 ______, 3 ______, 2 ______, 1 ______

TOTAL NUMBER OF WORDS ______ TOTAL NUMBER OF INDEX TERMS ______

Figure 5. Indexing Form Used in Performing the Content Analysis for the Rational Condensation Short-cut Method

TABLE 10

DEFINITION OF THE 23 QUANTITATIVE VARIABLES DERIVED FROM THE RATIONAL CONDENSATION INDEXING FORM

Number of Variable	Description of Variable
1	Weighted Frequency of Mention of MANAGEMENT FUNCTIONS
2	Weighted Frequency of Mention of LEADERSHIP AND DIRECTING
3	Weighted Frequency of Mention of ORGANIZATION AND STAFFING
4	Weighted Frequency of Mention of PLANNING-CONTROLLING
5	Weighted Frequency of Mention of REPRESENTATION
6	Weighted Frequency of Mention of SKILLS AND ABILITIES
7	Weighted Frequency of Mention of COMMUNICATION
8	Weighted Frequency of Mention of CONDUCT AND ATTITUDE
9	Weighted Frequency of Mention of COOPERATION AND RESPONSIVENESS
10	Weighted Frequency of Mention of CREATIVITY AND INITIATIVE
11	Weighted Frequency of Mention of ENDURANCE AND MOTIVATION
12	Weighted Frequency of Mention of INTELLECTUAL FUNCTIONING
13	Weighted Frequency of Mention of PROFESSIONAL AND TECHNICAL SKILLS
14	Weighted Frequency of Mention of PRODUCTIVITY AND ACHIEVEMENT
15	Weighted Frequency of Mention of RECOGNITION
16	Sum of Variables 1 through 15
17	Total Number of 5 Weights
18	Total Number of 4 Weights
19	Total Number of 3 Weights
20	Total Number of 2 Weights
21	Total Number of 1 Weights
22	Total Number of Index Terms Used
23	Total Number of Words in Narrative Text

Figure 6 portrays the indexing form that was developed for use with the statistically selected subset method. From this list of 15 statistically selected index terms, 21 quantitative variables were derived (see Table 11). The first 15 variables represent the *weighted* frequency of each of the statistically selected index terms used to index a particular section of narrative text using the original weighting scale shown in Table 7. The simple frequency of each of these 15 terms was not employed because the weighted frequency provides more information. Variable 16 is the sum of the 15 weighted frequencies.

Variable 17 represents the frequency count over the entire statistically selected subset indexing form for all 5 weights. Variable 18 represents a similar frequency count for all 3 weights, and Variable 19 represents a similar frequency count for all 2 weights. Variable 20 is the total number of index terms of the 15 available that were used to index the section of narrative text. Variable 21 is the total number of words in the section of narrative text that was indexed.

ID No.__________ Criterion Group______________ Section__________

Index Term		Weighted Frequency
MANAGEMENT FUNCTIONS______	____________	______
LEADERSHIP AND DIRECTING___	____________	______
PLANNING__________	____________	______
REPRESENTATION______	____________	______
STAFFING__________	____________	______
SKILLS AND ABILITIES______	____________	______
COMMUNICATION______	____________	______
PROFESSIONALISM______	____________	______
RESPONSIVENESS______	____________	______
TECHNICAL SKILLS______	____________	______
PRODUCTIVITY AND ACHIEVEMENT___	____________	______
DRIVE__________	____________	______
POTENTIAL__________	____________	______
REPUTE__________	____________	______
ASSET TO THE NAVY___	____________	______

FREQUENCY COUNTS: 5 ______, 3 ______, 2 ______,

TOTAL NUMBER OF WORDS ______ TOTAL NUMBER OF INDEX TERMS ______

Figure 6. Indexing Form Used in Performing the Content Analysis for the Statistically Selected Subset Short-cut Method

TABLE 11

DEFINITION OF THE 21 QUANTITATIVE VARIABLES DERIVED FROM THE STATISTICALLY SELECTED SUBSET INDEXING FORM

Number of Variable	Description of Variable
1	Weighted Frequency of Mention of MANAGEMENT FUNCTIONS
2	Weighted Frequency of Mention of LEADERSHIP AND DIRECTING
3	Weighted Frequency of Mention of PLANNING
4	Weighted Frequency of Mention of REPRESENTATION
5	Weighted Frequency of Mention of STAFFING
6	Weighted Frequency of Mention of SKILLS AND ABILITIES
7	Weighted Frequency of Mention of COMMUNICATION
8	Weighted Frequency of Mention of PROFESSIONALISM
9	Weighted Frequency of Mention of RESPONSIVENESS
10	Weighted Frequency of Mention of TECHNICAL SKILLS
11	Weighted Frequency of Mention of PRODUCTIVITY AND ACHIEVEMENT
12	Weighted Frequency of Mention of DRIVE
13	Weighted Frequency of Mention of POTENTIAL
14	Weighted Frequency of Mention of REPUTE
15	Weighted Frequency of Mention of ASSET TO THE NAVY
16	Sum of Variables 1 through 15
17	Total Number of 5 Weights
18	Total Number of 3 Weights
19	Total Number of 2 Weights
20	Total Number of Index Terms Used
21	Total Number of Words in Narrative Text

SECTION 4. COMPARISON OF THE SHORT-CUT INDEXING METHODS WITH THE ORIGINAL LENGTHY PROCEDURE

Profiles or vectors of the values for the 23 quantitative variables derived from the indexing form for the rational condensation short-cut method (see Table 10) were prepared for all of the Evaluation Reports contained in the pilot study, cross validation, and generalization samples. Separate profiles were compiled for the Evaluation and Justification Sections of each Evaluation Report. The transformed weighting scale shown in Table 9 was used, and a constant of 10 was added to the weighted frequency of Variables 1 through 15 in order to avoid the incidence of any negative input values in the subsequent statistical computations. All profiles were transformed to the new weighting scale, entered onto IBM coding forms, and keypunched. Card decks for each of the three samples were assembled in six parts: (1) Upper Criterion Group - Evaluation Section, (2) Middle Criterion Group - Evaluation Section, (3) Lower Criterion Group - Evaluation Section, (4) Upper Criterion Group - Justification Section, (5) Middle Criterion Group - Justification Section, and (6) Lower Criterion Group - Justification Section.

Profiles or vectors of the values for the 21 quantitative variables derived from the indexing form for the statistically selected subset short-cut method (see Table 11) were prepared in a similar manner, using the transformed weighting scale. Parallel card decks for the three samples were keypunched and assembled in six parts as described above for the rational condensation card decks.

The card decks corresponding to the two short-cut indexing methods were analyzed by Program BMD07M in the library of Biomedical Computer Programs[12] at the UCLA Health Sciences Computing Facility, just as the card decks for the original lengthy indexing procedure had been analyzed the year before.* Program BMD07M performs a multiple discriminant analysis in a stepwise manner. At each step one variable is entered into the set of discriminating variables or a variable is deleted if its F value becomes too low. At the option of the user, a classification matrix is computed and printed after those steps specified by the user. This option permits the user to determine if the classification process tends to converge to perfect classification or to maximize at some midway step and then diverge as more variables are added to the discriminant function.

Comparisons then were made among the three indexing methods for each step in the stepwise discriminant analysis procedure. These comparisons examined two aspects of the performance of each indexing method: (1) the variable that was selected at each step in the stepwise discriminant analysis procedure, and (2) the number of individuals in the particular sample being compared that were classified correctly into the criterion group to which they actually belonged. The remainder of this section presents a discussion of these comparisons. The results for each of the four occupational specialties represented in this study

---

*Computing assistance was obtained from the Health Sciences Computing Facility, UCLA, sponsored by NIH Special Research Resources Grant RR-3.

---AT's, BT's, CS's, and RM's---are presented separately, followed by a summary and discussion of the conclusions that can be drawn from the various comparisons that were made.

AT Comparison

AT Comparison - Evaluation Section. Table 12 shows a comparison of the performance of the three indexing procedures for the 144 pilot study AT's on the Evaluation Section of the Evaluation Report. At Step 1 the same variable was selected for all three indexing procedures---Total Number of 5 Weights (Excellent)---and the classification accuracy achieved by the three indexing procedures at Step 1 was approximately the same. The same variable---Total Number of 2 Weights (Poor)---was selected also for all three indexing procedures at Step 2. The classification accuracy dropped slightly for all three indexing procedures, but remained comparatively the same. Beginning at Step 3 there is a divergence among the three indexing procedures in the variable selected and some oscillation in the classification accuracy achieved. However, the following important clusters of variables were selected for at least two of the three indexing procedures between Steps 3 and 15: COMMUNICATION; PRODUCTIVITY AND ACHIEVEMENT; ORGANIZATION AND STAFFING/STAFFING; MANAGEMENT FUNCTIONS; COOPERATION/COOPERATION AND RESPONSIVENESS; REPRESENTATION; Total Number of Index Terms Used; and SERVICE MOTIVATION/ENDURANCE AND MOTIVATION/DRIVE.

At Step 14 the statistically selected subset indexing method achieved its best classification performance, correctly classifying 83 of the 144 pilot study AT's (58%). The best classification performance for the rational condensation indexing method was achieved at Step 18, with 80 of the 144 pilot study AT's (56%) being classified correctly. The stepwise discriminant analysis procedure continued to try to maximize the classification performance of the lengthy indexing method, using the greater number of available variables for this method. Finally, at Step 50 the best classification accuracy was achieved for the lengthy indexing procedure, 105 of the 144 pilot study AT's (73%). However, this superior classification accuracy for the lengthy indexing method compared to the two short-cut indexing methods was achieved because of the larger number of variables available to the stepwise discriminant analysis procedure for the lengthy method (67 compared to 23 for the rational condensation method and 21 for the statistically selected subset method). Perhaps a more meaningful comparison is the classification performance achieved by each indexing method at Step 15. At this step the lengthy method correctly classified 87 of the 144 pilot study AT's (60%) whereas the rational condensation method correctly classified 75 of the 144 pilot study AT's (52%). The statistically selected subset method had already reached its best classification performance at Step 14, correctly classifying 83 of the 144 pilot study AT's (58%). Even at Step 15 the lengthy indexing procedure shows a slight but definite advantage over the two short-cut indexing methods. Probably more interesting is the fact that the same variables were selected at Steps 1 and 2 for all three indexing methods. This is a significant finding since the variables selected early in the stepwise discriminant analysis procedure are the key variables in maximizing the ability of the discriminant function to differentiate membership in the three criterion groups---Upper, Middle, and Lower. From the statistical results presented in Table 12, it appears that the modifying adjectives and

TABLE 12

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY AT's (N=144) - EVALUATION SECTION

	Variable Selected	No. of Pilot Study AT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	68
Rational Condensation	Total Number of 5 (New 3) Weights*	68
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	67
Step 2:		
Lengthy Procedure	Total Number of 2 (New -1) Weights*	64
Rational Condensation	Total Number of 2 (New -1) Weights*	64
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	63
Step 3:		
Lengthy Procedure	wf of AWARDS AND PUNISHMENT	83
Rational Condensation	wf of COMMUNICATION	69
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	65
Step 4:		
Lengthy Procedure	f of COMMUNICATION	76
Rational Condensation	wf of ORGANIZATION AND STAFFING	72
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	80
Step 5:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	79
Rational Condensation	wf of CREATIVITY AND INITIATIVE	73
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	73
Step 6:		
Lengthy Procedure	f of COOPERATION	81
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	75
Stat. Selected Subset	wf of PROFESSIONALISM	76
Step 7:		
Lengthy Procedure	wf of MANAGEMENT FUNCTIONS	81
Rational Condensation	Sum of Variables 1 through 15	74
Stat. Selected Subset	wf of REPRESENTATION	76

* A 5 (New 3) Weight = Excellent; a 3 (New 1) Weight = Average; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 12 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY AT's (N=144) - EVALUATION SECTION

	Variable Selected	No. of Pilot Study AT's Classified Correctly
Step 8:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	89
Rational Condensation	Total Number of Index Terms Used	73
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	77
Step 9:		
Lengthy Procedure	wf of SERVICE MOTIVATION	88
Rational Condensation	wf of REPRESENTATION	71
Stat. Selected Subset	wf of COMMUNICATION	76
Step 10:		
Lengthy Procedure	Total Number of Index Terms Used	91
Rational Condensation	wf of SKILLS AND ABILITIES	69
Stat. Selected Subset	wf of ASSET TO THE NAVY	76
Step 11:		
Lengthy Procedure	wf of USE OF COMMUNICATION	93
Rational Condensation	wf of MANAGEMENT FUNCTIONS	73
Stat. Selected Subset	wf of STAFFING	75
Step 12:		
Lengthy Procedure	f of USE OF COMMUNICATION	89
Rational Condensation	wf of ENDURANCE AND MOTIVATION	73
Stat. Selected Subset	Total Number of Index Terms Used	82
Step 13:		
Lengthy Procedure	wf of REPRESENTATION	90
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	74
Stat. Selected Subset	wf of DRIVE	81
Step 14:		
Lengthy Procedure	wf of GROOMING AND ATTIRE	87
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	75
Stat. Selected Subset	wf of PLANNING	83*

---

* The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 12 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY AT's (N=144) - EVALUATION SECTION

	Variable Selected	No. of Pilot Study AT's Classified Correctly
Step 15:		
Lengthy Procedure	wf of DRIVE	87
Rational Condensation	Total Number of Words in Text	75
Step 16:		
Rational Condensation	Total Number of 4 (New 2) Weights*	78
Step 17:		
Rational Condensation	wf of CONDUCT AND ATTITUDE	77
Step 18:		
Rational Condensation	wf of LEADERSHIP AND DIRECTING	80**
Step 50:		
Lengthy Procedure	wf of ORGANIZATION	105**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	105 out of 144 (73%)
Rational Condensation	80 out of 144 (56%)
Stat. Selected Subset	83 out of 144 (58%)

---

* A 4 (New 2) Weight = Good.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

adverbs used by an evaluator to assess an individual are key factors in distinguishing between superior performance and less stellar achievements.

Do these findings also apply to the Evaluation Section for the cross validation AT's? Table 13 provides the answer. The results shown in Table 13 are less striking than those presented in Table 12. However, within the first five steps shown in Table 13, all three indexing procedures selected the same two variables dealing with weights, i.e., Total Number of 5 Weights (Excellent) and Total Number of 2 Weights (Poor). Other important clusters of variables selected in the first ten steps for all three of the indexing methods were COOPERATION/RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; LEADERSHIP AND DIRECTING; TECHNICAL SKILLS/PROFESSIONAL AND TECHNICAL SKILLS; and PRODUCTIVITY AND ACHIEVEMENT. Between Steps 8 and 15 the following additional clusters of variables were selected for two of the three indexing procedures: REPRESENTATION; SERVICE MOTIVATION/ENDURANCE AND MOTIVATION; Total Number of Words in Text; ORGANIZATION AND STAFFING/STAFFING; MANAGEMENT FUNCTIONS; and COMMUNICATION.

The classification performance of the three indexing procedures began at comparatively the same level in Table 13, but by the second step the statistically selected subset method began to fall behind and never recouped its losses while the rational condensation method managed to keep up reasonably well with the lengthy indexing procedure, even exceeding its classification performance at Steps 6, 7, and 8. At Step 16 the statistically selected subset indexing method achieved its best classification performance, correctly classifying 87 of the 138 cross validation AT's (63%). The rational condensation indexing method achieved its best classification performance at Step 18, correctly classifying 96 of the 138 cross validation AT's (70%). The lengthy indexing procedure, with its greater complement of variables, continued on to Step 48 where 110 of the 138 cross validation AT's were classified correctly (80%). Again, taking a reading at Step 15, one can see that the rational condensation method was fairly close to the original lengthy indexing procedure (65% accuracy compared to 68% accuracy), but the statistically selected subset method had lost ground (61%).

Since the pilot study and cross validation AT samples were drawn from data pools available for two contiguous years, it is possible that there were subtle differences in the composition of these two samples that made themselves evident in the results from the stepwise discriminant analysis procedure for each sample analyzed separately. Therefore, the pilot study and cross validation AT samples were combined into a single larger sample (N=282) and reanalyzed by the stepwise discriminant analysis procedure. Table 14 presents the results for the combined AT analysis on the Evaluation Section.

Now the importance of the variable, Total Number of 5 Weights (Excellent), becomes crystal clear, being selected initially for all three indexing procedures. The variable, Total Number of 2 Weights (Poor), was selected for all three indexing procedures within the first three steps. LEADERSHIP AND DIRECTING was selected for all three indexing procedures within the first four steps. TECHNICAL SKILLS/PROFESSIONAL AND TECHNICAL SKILLS plus MANAGEMENT FUNCTIONS were variables selected between Steps 4 and 6 for all three indexing procedures. The variable, Total Number of 3 Weights (Average), was selected by all three indexing procedures between Steps 2 and 12. Between Steps 6 and 15 four

TABLE 13

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION AT's (N=138) - EVALUATION SECTION

	Variable Selected	No. of Cross Valid. AT's Classified Correctly
Step 1:		
Lengthy Procedure	f of COOPERATION	58
Rational Condensation	Total Number of 5 (New 3) Weights*	61
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	59
Step 2:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	73
Rational Condensation	Total Number of 3 (New 1) Weights*	69
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	64
Step 3:		
Lengthy Procedure	f of TECHNICAL SKILLS	78
Rational Condensation	Total Number of 2 (New -1) Weights*	76
Stat. Selected Subset	Total Number of 5 (New 3) Weights *	70
Step 4:		
Lengthy Procedure	f of LEADERSHIP AND DIRECTING	85
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	78
Stat. Selected Subset	wf of TECHNICAL SKILLS	75
Step 5:		
Lengthy Procedure	Total Number of 2 (New -1) Weights*	84
Rational Condensation	wf of LEADERSHIP AND DIRECTING	80
Stat. Selected Subset	wf of RESPONSIVENESS	74
Step 6:		
Lengthy Procedure	wf of RESPONSIVENESS	82
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	86
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	68
Step 7:		
Lengthy Procedure	f of RESOURCEFULNESS	86
Rational Condensation	wf of CONDUCT AND ATTITUDE	89
Stat. Selected Subset	wf of PLANNING	71

* A 5 (New 3) Weight = Excellent; a 3 (New 1) Weight = Average; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 13 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION AT's (N=138) - EVALUATION SECTION

	Variable Selected	No. of Cross Valid. AT's Classified Correctly
Step 8:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	89
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	90
Stat. Selected Subset	wf of REPRESENTATION	76
Step 9:		
Lengthy Procedure	f of SERVICE MOTIVATION	95
Rational Condensation	wf of REPRESENTATION	89
Stat. Selected Subset	Total Number of Words in Text	81
Step 10:		
Lengthy Procedure	wf of TECHNICAL SKILLS	92
Rational Condensation	wf of ORGANIZATION AND STAFFING	89
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	80
Step 11:		
Lengthy Procedure	Total Number of 4 (New 2) Weights*	90
Rational Condensation	wf of MANAGEMENT FUNCTIONS	89
Stat. Selected Subset	wf of POTENTIAL	82
Step 12:		
Lengthy Procedure	f of AWARDS AND PUNISHMENT	92
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	87
Stat. Selected Subset	wf of ASSET TO THE NAVY	81
Step 13:		
Lengthy Procedure	f of PROFESSIONALISM	92
Rational Condensation	wf of COMMUNICATION	86
Stat. Selected Subset	wf of STAFFING	80
Step 14:		
Lengthy Procedure	Total Number of Index Terms Used	93
Rational Condensation	wf of ENDURANCE AND MOTIVATION	90
Stat. Selected Subset	wf of COMMUNICATION	84

* A 4 (New 2) Weight = Good.

(Continued)

TABLE 13 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION AT's (N=138) - EVALUATION SECTION

	Variable Selected	No. of Cross Valid. AT's Classified Correctly
Step 15:		
Lengthy Procedure	Total Number of Words in Text	94
Rational Condensation	wf of SKILLS AND ABILITIES	90
Stat. Selected Subset	wf of PROFESSIONALISM	84
Step 16:		
Rational Condensation	Total Number of 4 (New 2) Weights*	91
Stat. Selected Subset	wf of REPUTE	87**
Step 17:		
Rational Condensation	wf of RECOGNITION	94
Step 18:		
Rational Condensation	wf of CREATIVITY AND INITIATIVE	96**
Step 48:		
Lengthy Procedure	f of INITIATIVE	110**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	110 out of 138 (80%)
Rational Condensation	96 out of 138 (70%)
Stat. Selected Subset	87 out of 138 (63%)

---

* A 4 (New 2) Weight = Good.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

TABLE 14

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED AT SAMPLES (N=282) - EVALUATION SECTION

	Variable Selected	No. of Combined AT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	129
Rational Condensation	Total Number of 5 (New 3) Weights*	129
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	130
Step 2:		
Lengthy Procedure	f of LEADERSHIP AND DIRECTING	142
Rational Condensation	Total Number of 3 (New 1) Weights*	138
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	121
Step 3:		
Lengthy Procedure	Total Number of 2 (New -1) Weights*	151
Rational Condensation	Total Number of 2 (New -1) Weights*	143
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	147
Step 4:		
Lengthy Procedure	wf of TECHNICAL SKILLS	151
Rational Condensation	wf of LEADERSHIP AND DIRECTING	162
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	142
Step 5:		
Lengthy Procedure	Total Number of 3 (New 1) Weights*	158
Rational Condensation	wf of MANAGEMENT FUNCTIONS	155
Stat. Selected Subset	wf of TECHNICAL SKILLS	142
Step 6:		
Lengthy Procedure	wf of MANAGEMENT FUNCTIONS	165
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	166**
Stat. Selected Subset	wf of PROFESSIONALISM	143
Step 7:		
Lengthy Procedure	wf of RESPONSIVENESS	162
Stat. Selected Subset	wf of REPRESENTATION	146

* A 5 (New 3) Weight = Excellent; a 3 (New 1) Weight = Average; a 2 (New -1) Weight = Poor.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 14 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED AT SAMPLES (N=282) - EVALUATION SECTION

	Variable Selected	No. of Combined AT's Classified Correctly
Step 8:		
Lengthy Procedure	f of COMMUNICATION	160
Stat. Selected Subset	wf of RESPONSIVENESS	145
Step 9:		
Lengthy Procedure	f of RESOURCEFULNESS	165
Stat. Selected Subset	wf of POTENTIAL	147
Step 10:		
Lengthy Procedure	f of PROFESSIONALISM	162
Stat. Selected Subset	wf of COMMUNICATION	147
Step 11:		
Lengthy Procedure	f of DRIVE	157
Stat. Selected Subset	Total Number of Words in Text	152
Step 12:		
Lengthy Procedure	wf of LEADERSHIP AND DIRECTING	157
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	152
Step 13:		
Lengthy Procedure	f of AWARDS AND PUNISHMENT	162
Stat. Selected Subset	wf of DRIVE	150
Step 14:		
Lengthy Procedure	wf of POTENTIAL	163
Stat. Selected Subset	wf of STAFFING	151
Step 15:		
Lengthy Procedure	wf of INTELLECTUAL FUNCTIONING	160
Stat. Selected Subset	Total Number of Index Terms Used	150
Step 16:		
Stat. Selected Subset	Sum of Variables 1 through 15	156

* A 3 (New 1) Weight = Average.

(Continued)

TABLE 14 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED AT SAMPLES (N=282) - EVALUATION SECTION

	Variable Selected	No. of Combined AT's Classified Correctly
Step 17:		
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	152
Step 18:		
Stat. Selected Subset	wf of ASSET TO THE NAVY	151
Step 19:		
Stat. Selected Subset	wf of REPUTE	152
Step 20:		
Lengthy Procedure	wf of CONDUCT, INTEGRITY, AND PRIDE	164
Stat. Selected Subset	wf of SKILLS AND ABILITIES	152
Step 21:		
Stat. Selected Subset	wf of PLANNING	159*
Step 46:		
Lengthy Procedure	wf of DRIVE	184*

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	184 out of 282 (65%)
Rational Condensation	166 out of 282 (59%)
Stat. Selected Subset	159 out of 282 (56%)

---

* The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

variables that assumed importance were RESPONSIVENESS, COMMUNICATION, POTENTIAL, and DRIVE, being selected by at least two of the three indexing procedures.

At Step 6 the rational condensation indexing method achieved its best classification performance, correctly classifying 166 of the 282 combined AT's (59%). The statistically selected subset indexing method lagged behind but continued on to Step 21 where it achieved its best classification performance, correctly classifying 159 of the 282 combined AT's (56%). The lengthy indexing procedure finally reached its maximum classification performance at Step 46 where it correctly classified 184 of the 282 combined AT's (65%).

AT Comparison - Justification Section. Table 15 shows a comparison of the performance of the three indexing procedures for the 144 pilot study AT's on the Justification Section of the Evaluation Report. For all three indexing procedures, the first variable selected was Total Number of Index Terms Used, with the classification accuracy for the three methods being approximately the same initially. The lengthy procedure and the rational condensation method paralleled each other at Step 2, with Total Number of 5 Weights (Excellent) being selected for both procedures. The three procedures then go their more or less independent ways until Step 15, but certain important clusters of variables were selected by at least two of the indexing methods up to this point. These clusters were REPUTE/POTENTIAL/RECOGNITION; INTELLECTUAL FUNCTIONING; PRODUCTIVITY AND ACHIEVEMENT; Total Number of 3 Weights (Average); COOPERATION AND RESPONSIVENESS/RESPONSIVENESS; ORGANIZATION AND STAFFING/STAFFING; ENDURANCE AND MOTIVATION/ENDURANCE/DRIVE/SERVICE MOTIVATION; SKILLS AND ABILITIES; PLANNING/CONTROLLING/PLANNING-CONTROLLING; and sum of either the simple or weighted frequencies of the available set of variables for a particular indexing method.

At Step 15 in Table 15 the statistically selected subset indexing method achieved its best classification performance, correctly classifying 116 of the 144 pilot study AT's (81%). The lengthy indexing procedure was only slightly better at Step 15, correctly classifying 117 of the 144 pilot study AT's (81%). At Step 15 the rational condensation indexing method was somewhat inferior to the other two indexing procedures, correctly classifying 109 of the 144 pilot study AT's (76%). By Step 18 the rational condensation method achieved its best classification performance, correctly classifying 110 of the 144 pilot study AT's (76%). The stepwise discriminant analysis procedure continued on to Step 58 where it finally achieved a best classification performance of 137 individuals in the 144-case pilot study AT sample (95%) for the lengthy indexing procedure.

Table 16 presents the Justification Section results for the 138 cross validation AT's. As with the pilot study AT's on the Justification Section, the variable, Total Number of Index Terms Used, was selected initially for all three indexing procedures, with the classification accuracy for the three methods being approximately the same. At Step 2 TECHNICAL SKILLS was selected for the original lengthy indexing procedure and for the statistically selected subset short-cut method, and the new index term, PROFESSIONAL AND TECHNICAL SKILLS, was selected for the rational condensation short-cut indexing method; again, the classification accuracy of the three methods was approximately the same. At Step 3 PRODUCTIVITY AND ACHIEVEMENT was the variable selected for both the lengthy indexing procedure and for the rational condensation method,

TABLE 15

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY AT's (N=144) - JUSTIFICATION SECTION

	Variable Selected	No. of Pilot Study AT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	95
Rational Condensation	Total Number of Index Terms Used	96
Stat. Selected Subset	Total Number of Index Terms Used	99
Step 2:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	104
Rational Condensation	Total Number of 5 (New 3) Weights*	106
Stat. Selected Subset	wf of REPUTE	99
Step 3:		
Lengthy Procedure	f of CONDUCT, INTEGRITY, AND PRIDE	105
Rational Condensation	Total Number of 4 (New 2) Weights*	107
Stat. Selected Subset	Total Number of Words in Text	100
Step 4:		
Lengthy Procedure	f of INTELLECTUAL FUNCTIONING	103
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	108
Stat. Selected Subset	wf of PROFESSIONALISM	103
Step 5:		
Lengthy Procedure	wf of INTELLECTUAL FUNCTIONING	102
Rational Condensation	Total Number of 3 (New 1) Weights*	105
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	109
Step 6:		
Lengthy Procedure	wf of GROOMING AND ATTIRE	104
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	106
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	111
Step 7:		
Lengthy Procedure	f of TECHNICAL SKILLS	108
Rational Condensation	wf of ORGANIZATION AND STAFFING	109
Stat. Selected Subset	wf of ASSET TO THE NAVY	107

* A 5 (New 3) Weight = Excellent; a 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average.

(Continued)

TABLE 15 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY AT's (N=144) - JUSTIFICATION SECTION

	Variable Selected	No. of Pilot Study AT's Classified Correctly
Step 8:		
Lengthy Procedure	f of REPUTE	110
Rational Condensation	wf of ENDURANCE AND MOTIVATION	109
Stat. Selected Subset	wf of STAFFING	109
Step 9:		
Lengthy Procedure	wf of FLEXIBILITY	111
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	106
Stat. Selected Subset	wf of POTENTIAL	109
Step 10:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	112
Rational Condensation	wf of MANAGEMENT FUNCTIONS	106
Stat. Selected Subset	wf of SKILLS AND ABILITIES	109
Step 11:		
Lengthy Procedure	Sum of Variables 1 through 29	110
Rational Condensation	wf of REPRESENTATION	106
Stat. Selected Subset	Sum of Variables 1 through 15	111
Step 12:		
Lengthy Procedure	wf of ENDURANCE	112
Rational Condensation	wf of PLANNING-CONTROLLING	106
Stat. Selected Subset	wf of DRIVE	111
Step 13:		
Lengthy Procedure	wf of SERVICE MOTIVATION	114
Rational Condensation	wf of RECOGNITION	108
Stat. Selected Subset	wf of PLANNING	110
Step 14:		
Lengthy Procedure	wf of CONTROLLING	117
Rational Condensation	Total Number of 1 (New -2) Weights*	108
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	115

* A 1 (New -2) Weight = Poorest.

(Continued)

TABLE 15 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY AT's (N=144) - JUSTIFICATION SECTION

	Variable Selected	No. of Pilot Study AT's Classified Correctly
Step 15:		
Lengthy Procedure	f of ENDURANCE	117
Rational Condensation	wf of SKILLS AND ABILITIES	109
Stat. Selected Subset	wf of RESPONSIVENESS	116*
Step 16:		
Rational Condensation	wf of CREATIVITY AND INITIATIVE	109
Step 17:		
Rational Condensation	wf of LEADERSHIP AND DIRECTING	109
Step 18:		
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	110*
Step 58:		
Lengthy Procedure	f of POTENTIAL	137*

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	137 out of 144 (95%)
Rational Condensation	110 out of 144 (76%)
Stat. Selected Subset	116 out of 144 (81%)

---

* The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

TABLE 16

COMPARISON OF THE VARIABLES SELECTED AY EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION AT's (N=138) - JUSTIFICATION SECTION

	Variable Selected	No. of Cross Valid. AT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	88
Rational Condensation	Total Number of Index Terms Used	88
Stat. Selected Subset	Total Number of Index Terms Used	85
Step 2:		
Lengthy Procedure	wf of TECHNICAL SKILLS	97
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	98
Stat. Selected Subset	wf of TECHNICAL SKILLS	96
Step 3:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	98
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	100
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	96
Step 4:		
Lengthy Procedure	wf of SKILLS AND ABILITIES	102
Rational Condensation	wf of SKILLS AND ABILITIES	104
Stat. Selected Subset	wf of SKILLS AND ABILITIES	101
Step 5:		
Lengthy Procedure	f of COMMUNICATION	101
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	108
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	100
Step 6:		
Lengthy Procedure	wf of ORGANIZATION	100
Rational Condensation	wf of ORGANIZATION AND STAFFING	104
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	102
Step 7:		
Lengthy Procedure	wf of ASSET TO THE NAVY	102
Rational Condensation	wf of MANAGEMENT FUNCTIONS	100
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	101

---

* A 3 (New 1) Weight = Average.

(Continued)

TABLE 16 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION AT's (N=138) - JUSTIFICATION SECTION

	Variable Selected	No. of Cross Valid. AT's Classified Correctly
Step 8:		
Lengthy Procedure	wf of REPUTE	104
Rational Condensation	wf of ENDURANCE AND MOTIVATION	108
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	103
Step 9:		
Lengthy Procedure	wf of AWARDS AND PUNISHMENT	106
Rational Condensation	Total Number of 4 (New 2) Weights*	104
Stat. Selected Subset	wf of ASSET TO THE NAVY	103
Step 10:		
Lengthy Procedure	f of INTELLECTUAL FUNCTIONING	108
Rational Condensation	Total Number of Words in Text	109
Stat. Selected Subset	wf of REPUTE	102
Step 11:		
Lengthy Procedure	wf of POTENTIAL	109
Rational Condensation	wf of RECOGNITION	109
Stat. Selected Subset	wf of REPRESENTATION	104
Step 12:		
Lengthy Procedure	f of REPRESENTATION	107
Rational Condensation	wf of LEADERSHIP AND DIRECTING	108
Stat. Selected Subset	wf of POTENTIAL	103
Step 13:		
Lengthy Procedure	Total Number of Words in Text	108
Rational Condensation	Total Number of 2 (New -1) Weights*	107
Stat. Selected Subset	wf of COMMUNICATION	103
Step 14:		
Lengthy Procedure	wf of RESPONSIVENESS	111
Rational Condensation	Total Number of 3 (New 1) Weights*	109
Stat. Selected Subset	Total Number of Words in Text	103

* A 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 16 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION AT's (N=138) - JUSTIFICATION SECTION

	Variable Selected	No. of Cross Valid. AT's Classified Correctly
Step 15:		
Lengthy Procedure	f of RESPONSIVENESS	113
Rational Condensation	wf of REPRESENTATION	108
Stat. Selected Subset	wf of PLANNING	101
Step 16:		
Rational Condensation	wf of COMMUNICATION	110**
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	107
Step 17:		
Stat. Selected Subset	Sum of Variables 1 through 15	107
Step 18:		
Stat. Selected Subset	wf of PROFESSIONALISM	109**
Step 46:		
Lengthy Procedure	wf of COMMUNICATION	129**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	129 out of 138 (93%)
Rational Condensation	110 out of 138 (80%)
Stat. Selected Subset	109 out of 138 (79%)

---

* A 5 (New 3) Weight = Excellent.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

with this same variable being selected at Step 5 for the statistically selected subset method. At Step 4 the variable SKILLS AND ABILITIES was selected for all three indexing procedures After Step 4 in Table 16 the three indexing procedures go their separate ways until Step 15, but certain important clusters of variables were selected for at least two of the three indexing methods. These clusters were COMMUNICATION; INTELLECTUAL FUNCTIONING; ORGANIZATION/ORGANIZATION AND STAFFING; Total Number of 3 Weights (Average); MANAGEMENT FUNCTIONS; LEADERSHIP AND DIRECTING; ASSET TO THE NAVY/REPUTE/AWARDS AND PUNISHMENT/POTENTIAL/RECOGNITION; Total Number of 2 Weights (Poor); Total Number of Words in Text; REPRESENTATION; and RESPONSIVENESS.

At Step 16 the rational condensation indexing method achieved its best classification performance, correctly classifying 110 of the 138 cross validation AT's (80%). At Step 18 the statistically selected subset indexing method reached its best classification performance, correctly classifying 109 of the 138 cross validation AT's (79%). Finally, at Step 46 the lengthy indexing procedure attained its best classification performance, correctly classifying 129 of the 138 cross validation AT's (93%). However, a reading at Step 15 shows that the lengthy indexing procedure was only slightly superior to the two short-cut methods (82% classification accuracy for the lengthy procedure compared to 78% classification accuracy for the rational condensation method and 73% classification accuracy for the statistically selected subset method).

A combination of the pilot study and cross validation AT samples also was reanalyzed for the Justification Section. The results of this analysis are presented in Table 17. As expected, the variable, Total Number of Index Terms Used, was selected first for all three indexing procedures just as this variable universally was selected first when these two samples were analyzed separately. In Table 17 the initial classification accuracy for the three methods was approximately the same. At Step 2 in the combined AT analysis for the Justification Section, the sum of the weighted frequencies of the available set of variables for a particular indexing procedure was the variable selected for all three indexing methods. At Step 3 PRODUCTIVITY AND ACHIEVEMENT was selected for all three indexing methods. From Step 4 through 15 in Table 17 certain important clusters of variables were selected for at least two of the three indexing procedures. These clusters were PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS/PROFESSIONALISM; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; STAFFING/ORGANIZATION AND STAFFING; INTELLECTUAL FUNCTIONING; ENDURANCE AND MOTIVATION/DRIVE; MANAGEMENT FUNCTIONS; REPUTE; Total Number of 2 Weights (Poor); COMMUNICATION; REPRESENTATION; and Total Number of Words in Text. At Step 15 the rational condensation indexing method was slightly superior to the lengthy indexing procedure (77% classification accuracy compared to 76% classification accuracy), but the statistically selected subset indexing method had already reached its best classification performance at Step 12, correctly classifying 214 of the 282 combined AT's (76%). At Step 17 the rational condensation method achieved its best classification performance, correctly classifying 217 of the 282 combined AT's (77%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 46 where this procedure's best classification performance was achieved, correctly classifying 230 of the 282 combined AT's (82%).

TABLE 17

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED AT SAMPLES (N=282) - JUSTIFICATION SECTION

	Variable Selected	No. of Combined AT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	184
Rational Condensation	Total Number of Index Terms Used	182
Stat. Selected Subset	Total Number of Index Terms Used	181
Step 2:		
Lengthy Procedure	Sum of Variables 31 through 59	189
Rational Condensation	Sum of Variables 1 through 15	192
Stat. Selected Subset	Sum of Variables 1 through 15	185
Step 3:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	202
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	199
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	195
Step 4:		
Lengthy Procedure	Sum of Variables 1 through 29	198
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	202
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	197
Step 5:		
Lengthy Procedure	wf of TECHNICAL SKILLS	199
Rational Condensation	wf of SKILLS AND ABILITIES	206
Stat. Selected Subset	wf of SKILLS AND ABILITIES	191
Step 6:		
Lengthy Procedure	wf of STAFFING	203
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	205
Stat. Selected Subset	wf of TECHNICAL SKILLS	199
Step 7:		
Lengthy Procedure	wf of SKILLS AND ABILITIES	203
Rational Condensation	wf of ENDURANCE AND MOTIVATION	210
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	204

* A 3 (New 1) Weight = Average.

(Continued)

TABLE 17 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED AT SAMPLES (N=282) - JUSTIFICATION SECTION

	Variable Selected	No. of Combined AT's Classified Correctly
Step 8:		
Lengthy Procedure	f of REPUTE	203
Rational Condensation	Total Number of 3 (New 1) Weights*	206
Stat. Selected Subset	wf of REPUTE	203
Step 9:		
Lengthy Procedure	wf of DRIVE	208
Rational Condensation	wf of ORGANIZATION AND STAFFING	209
Stat. Selected Subset	wf of ASSET TO THE NAVY	209
Step 10:		
Lengthy Procedure	wf of PROFESSIONALISM	209
Rational Condensation	wf of MANAGEMENT FUNCTIONS	211
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	210
Step 11:		
Lengthy Procedure	f of INTELLECTUAL FUNCTIONING	210
Rational Condensation	Total Number of 2 (New -1) Weights*	211
Stat. Selected Subset	wf of COMMUNICATION	212
Step 12:		
Lengthy Procedure	f of AWARDS AND PUNISHMENT	212
Rational Condensation	wf of REPRESENTATION	212
Stat. Selected Subset	wf of STAFFING	214**
Step 13:		
Lengthy Procedure	Total Number of 3 (New 1) Weights*	211
Rational Condensation	Total Number of Words in Text	211
Step 14:		
Lengthy Procedure	wf of REPRESENTATION	212
Rational Condensation	wf of COMMUNICATION	215

* A 3 (New 1) Weight = Average; a 2 (New -1) Weight = Poor.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 17 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED AT SAMPLES (N=282) - JUSTIFICATION SECTION

	Variable Selected	No. of Combined AT's Classified Correctly
Step 15:		
Lengthy Procedure	Total Number of Words in Text	213
Rational Condensation	Total Number of 1 (New -2) Weights*	216
Step 16:		
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	214
Step 17:		
Rational Condensation	wf of LEADERSHIP AND DIRECTING	217**
Step 46:		
Lengthy Procedure	Total Number of 4 (New 2) Weights*	230**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	230 out of 282 (82%)
Rational Condensation	217 out of 282 (77%)
Stat. Selected Subset	214 out of 282 (76%)

---

* A 4 (New 2) Weight = Good; a 1 (New -2) Weight = Poorest.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

BT Comparison

BT Comparison - Evaluation Section. Table 18 shows a comparison of the performance of the three indexing procedures for the 80 pilot study BT's on the Evaluation Section of the Evaluation Report. At Step 1 the same variable was selected for all three indexing procedures---Total Number of 5 Weights (Excellent)---and the classification accuracy achieved by the three indexing procedures at Step 1 was approximately the same. The variable, Total Number of 5 Weights, was also the variable selected initially by all three indexing procedures for the pilot study AT's (see Table 12). At Step 2 the variable, Total Number of 2 Weights (Poor), was selected for both the lengthy indexing procedure and the rational condensation short-cut method, and at Step 7 this same variable was selected for the statistically selected subset method. The discriminating role of these two variables dealing with weights assumes importance with the BT's just as these variables did with the AT's in the analysis of the Evaluation Section of the pilot study sample. After Step 3 there is a divergence among the three indexing procedures in the variables selected, but certain clusters of variables were selected for at least two of the three indexing procedures---TECHNICAL SKILLS; LEADERSHIP AND DIRECTING; COMMUNICATION; RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; PROFESSIONALISM; MANAGEMENT FUNCTIONS; SKILLS AND ABILITIES; RECOGNITION/POTENTIAL/ASSET TO THE NAVY; and ORGANIZATION AND STAFFING/STAFFING.

At Step 12 the statistically selected subset method achieved its best classification performance, correctly classifying 57 of the 80 pilot study BT's (71%). The best classification performance for the rational condensation indexing method was achieved at Step 16, with 57 of the 80 pilot study BT's (71%) also being classified correctly. The stepwise discriminant analysis procedure continued to try to maximize the classification performance of the lengthy indexing method, using the greater number of available variables for this method. Finally, at Step 50 the best classification accuracy was achieved for the lengthy indexing procedure, 77 of the 80 pilot study BT's (96%). Even at Step 15 the lengthy indexing procedure showed a definite advantage over the two short-cut indexing methods (80% classification accuracy for the lengthy procedure compared to 68% classification accuracy for the rational condensation method and 71% maximum classification accuracy achieved at Step 12 for the statistically selected subset method). For the pilot study BT's as for the pilot study AT's, once again the modifying adjectives and adverbs used by an evaluator to assess an individual appear to be the key discriminating variables for the Evaluation Section of the Evaluation Report in distinguishing between superlative chief petty officers and their slightly less qualified colleagues.

The results for the comparison of the three indexing procedures on the Evaluation Section for the 84 cross validation BT's is shown in Table 19 where it can be seen that once again the variable, Total Number of 5 Weights (Excellent), was the first variable selected for all three indexing procedures. At Step 2 the variable, Total Number of Index Terms Used, was selected for all three indexing procedures. Other important clusters of variables selected in the first 15 steps for at least two of the three indexing procedures were sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of 2 Weights (Poor); RESOURCEFULNESS/CREATIVITY AND INITIATIVE; REPUTE/AWARDS AND PUNISHMENT/ASSET TO THE NAVY/RECOGNITION;

TABLE 18

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY BT's (N=80) - EVALUATION SECTION

	Variable Selected	No. of Pilot Study BT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	34
Rational Condensation	Total Number of 5 (New 3) Weights*	34
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	32
Step 2:		
Lengthy Procedure	Total Number of 2 (New -1) Weights*	45
Rational Condensation	Total Number of 2 (New -1) Weights*	45
Stat. Selected Subset	wf of TECHNICAL SKILLS	33
Step 3:		
Lengthy Procedure	f of LEADERSHIP AND DIRECTING	47
Rational Condensation	wf of COMMUNICATION	45
Stat. Selected Subset	wf of RESPONSIVENESS	42
Step 4:		
Lengthy Procedure	wf of PROFESSIONALISM	48
Rational Condensation	wf of MANAGEMENT FUNCTIONS	48
Stat. Selected Subset	wf of PROFESSIONALISM	45
Step 5:		
Lengthy Procedure	f of RESPONSIVENESS	51
Rational Condensation	Total Number of 1 (New -2) Weights*	49
Stat. Selected Subset	wf of SKILLS AND ABILITIES	46
Step 6:		
Lengthy Procedure	wf of GROOMING AND ATTIRE	53
Rational Condensation	wf of LEADERSHIP AND DIRECTING	48
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	49
Step 7:		
Lengthy Procedure	wf of TECHNICAL SKILLS	54
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	48
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	48

* A 5 (New 3) Weight = Excellent; a 2 (New -1) Weight = Poor; a 1 (New -2) Weight = Poorest.

(Continued)

TABLE 18 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY BT's (N=80) - EVALUATION SECTION

	Variable Selected	No. of Pilot Study BT's Classified Correctly
Step 8:		
Lengthy Procedure	wf of CONTROLLING	56
Rational Condensation	wf of SKILLS AND ABILITIES	50
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	51
Step 9:		
Lengthy Procedure	f of SKILLS AND ABILITIES	57
Rational Condensation	wf of RECOGNITION	51
Stat. Selected Subset	wf of COMMUNICATION	53
Step 10:		
Lengthy Procedure	wf of RELIABILITY AND DEPENDABILITY	56
Rational Condensation	wf of ENDURANCE AND MOTIVATION	48
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	54
Step 11:		
Lengthy Procedure	wf of MANAGEMENT FUNCTIONS	57
Rational Condensation	wf of ORGANIZATION AND STAFFING	49
Stat. Selected Subset	wf of POTENTIAL	54
Step 12:		
Lengthy Procedure	wf of USE OF COMMUNICATION	60
Rational Condensation	Total Number of 4 (New 2) Weights*	51
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	57**
Step 13:		
Lengthy Procedure	wf of ASSET TO THE NAVY	60
Rational Condensation	Total Number of Words in Text	51
Step 14:		
Lengthy Procedure	f of MANAGEMENT FUNCTIONS	64
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	53

* A 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 18 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY BT's (N=80) - EVALUATION SECTION

	Variable Selected	No. of Pilot Study BT's Classified Correctly
Step 15:		
Lengthy Procedure	wf of STAFFING	64
Rational Condensation	Total Number of Index Terms Used	54
Step 16:		
Rational Condensation	Total Number of 3 (New 1) Weights*	57**
Step 50:		
Lengthy Procedure	wf of POTENTIAL	77**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	77 out of 80 (96%)
Rational Condensation	57 out of 80 (71%)
Stat. Selected Subset	57 out of 80 (71%)

---

* A 3 (New 1) Weight = Average.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

TABLE 19

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION BT's (N=84) - EVALUATION SECTION

	Variable Selected	No. of Cross Valid. BT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	39
Rational Condensation	Total Number of 5 (New 3) Weights*	39
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	33
Step 2:		
Lengthy Procedure	Total Number of Index Terms Used	52
Rational Condensation	Total Number of Index Terms Used	50
Stat. Selected Subset	Total Number of Index Terms Used	37
Step 3:		
Lengthy Procedure	Sum of Variables 31 through 59	53
Rational Condensation	Total Number of 2 (New -1) Weights*	50
Stat. Selected Subset	Sum of Variables 1 through 15	42
Step 4:		
Lengthy Procedure	wf of RESOURCEFULNESS	53
Rational Condensation	Sum of Variables 1 through 15	50
Stat. Selected Subset	wf of PROFESSIONALISM	49
Step 5:		
Lengthy Procedure	f of COOPERATION	57
Rational Condensation	wf of CREATIVITY AND INITIATIVE	54
Stat. Selected Subset	wf of REPUTE	51
Step 6:		
Lengthy Procedure	f of RESOURCEFULNESS	58
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	53
Stat. Selected Subset	wf of SKILLS AND ABILITIES	55
Step 7:		
Lengthy Procedure	Total Number of Words in Text	60
Rational Condensation	wf of SKILLS AND ABILITIES	55
Stat. Selected Subset	wf of STAFFING	55

---

* A 5 (New 3) Weight = Excellent; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 19 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION BT's (N=84) - EVALUATION SECTION

	Variable Selected	No. of Cross Valid. BT's Classified Correctly
Step 8:		
Lengthy Procedure	f of AWARDS AND PUNISHMENT	58
Rational Condensation	wf of ORGANIZATION AND STAFFING	55
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	55
Step 9:		
Lengthy Procedure	f of ORGANIZATION	61
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	59
Stat. Selected Subset	wf of ASSET TO THE NAVY	58
Step 10:		
Lengthy Procedure	f of TECHNICAL SKILLS	62
Rational Condensation	wf of RECOGNITION	59
Stat. Selected Subset	wf of RESPONSIVENESS	58
Step 11:		
Lengthy Procedure	f of ASSET TO THE NAVY	62
Rational Condensation	Total Number of 1 (New -2) Weights*	59
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	58
Step 12:		
Lengthy Procedure	wf of ASSET TO THE NAVY	64
Rational Condensation	Total Number of Words in Text	62
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	59
Step 13:		
Lengthy Procedure	wf of FLEXIBILITY	64
Rational Condensation	wf of LEADERSHIP AND DIRECTING	61
Stat. Selected Subset	wf of COMMUNICATION	62
Step 14:		
Lengthy Procedure	wf of REPUTE	62
Rational Condensation	wf of MANAGEMENT FUNCTIONS	63**
Stat. Selected Subset	wf of REPRESENTATION	62

* A 2 (New -1) Weight = Poor; a 1 (New -2) Weight = Poorest.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 19 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION BT's (N=84) - EVALUATION SECTION

	Variable Selected	No. of Cross Valid. BT's Classified Correctly
Step 15:		
Lengthy Procedure	f of STAFFING	65
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	61
Step 16:		
Stat. Selected Subset	wf of POTENTIAL	60
Step 17:		
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	62
Step 18:		
Stat. Selected Subset	Total Number of Words in Text	61
Step 19:		
Stat. Selected Subset	wf of TECHNICAL SKILLS	61
Step 20:		
Lengthy Procedure	f of PROFESSIONALISM	69
Stat. Selected Subset	wf of DRIVE	60
Step 21:		
Stat. Selected Subset	wf of PLANNING	64**
Step 62:		
Lengthy Procedure	f of COOPERATION	82**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	82 out of 84 (98%)
Rational Condensation	63 out of 84 (75%)
Stat. Selected Subset	64 out of 84 (76%)

---

[*] A 3 (New 1) Weight = Average.

[**] The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

PRODUCTIVITY AND ACHIEVEMENT; SKILLS AND ABILITIES; Total Number of Words in Text; STAFFING/ORGANIZATION AND STAFFING/ORGANIZATION; and MANAGEMENT FUNCTIONS.

The classification performance of the lengthy indexing procedure and the rational condensation indexing method began at the same level in Table 19, with the statistically selected subset method lagging behind somewhat. The statistically selected subset method continued to trail the other two indexing procedures in classification performance until Step 6 where all three procedures achieved approximately the same classification accuracy. At Step 14 the three methods are almost identical in their classification performance, and it is at this step that the rational condensation method reached its best classification performance, correctly classifying 63 of the 84 cross validation BT's (75%). The statistically selected subset method achieved its best classification performance at Step 21, correctly classifying 64 of the 84 cross validation BT's (76%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 62, trying to maximize its classification performance with the greater number of variables available to it, finally correctly classifying 82 of the 84 cross validation BT's (98%).

As with the AT samples discussed earlier, the pilot study and cross validation BT samples were combined into a single larger sample (N=164) and reanalyzed by the stepwise discriminant analysis procedure. Table 20 presents the results for the combined BT analysis on the Evaluation Section.

Once again the importance of the variable, Total Number of 5 Weights (Excellent), becomes unmistakably apparent, being selected initially for all three indexing procedures. At either Step 2 or Step 3 the two variables, Total Number of 2 Weights (Poor) and Total Number of Index Terms Used, were selected for all three indexing procedures. Between Steps 4 and 15 the following clusters of variables were selected for at least two of the three indexing procedures: MANAGEMENT FUNCTIONS; COMMUNICATION; PROFESSIONALISM; RECOGNITION/REPUTE/ASSET TO THE NAVY/POTENTIAL; SKILLS AND ABILITIES; RESOURCEFULNESS/CREATIVITY AND INITIATIVE; REPRESENTATION; ORGANIZATION/STAFFING/ORGANIZATION AND STAFFING; and PRODUCTIVITY AND ACHIEVEMENT.

At Step 15 the lengthy indexing procedure held a slight edge over the two short-cut indexing methods (63% classification accuracy for the lengthy procedure compared to 60% classification accuracy for the rational condensation method and 57% classification accuracy for the statistically selected subset method). At Step 16 the rational condensation method attained its best classification performance, correctly classifying 100 of the 164 combined BT's (61%). At Step 17 the statistically selected subset method attained its best classification performance, correctly classifying 106 of the 164 combined BT's (65%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 54, finally achieving a classification accuracy of 123 individuals in the 164-case combined BT sample (75%).

BT Comparison - Justification Section. Table 21 shows a comparison of the performance of the three indexing procedures for the 80 pilot study BT's on the Justification Section of the Evaluation Report. For all three indexing procedures, the first variable selected was Total Number of Index Terms Used, with the classification accuracy for the three methods being approximately the same initially. The variable, Total Number of Index Terms Used,

TABLE 20

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED BT SAMPLES (N=164) - EVALUATION SECTION

	Variable Selected	No. of Combined BT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	73
Rational Condensation	Total Number of 5 (New 3) Weights*	73
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	65
Step 2:		
Lengthy Procedure	Total Number of 2 (New -1) Weights*	84
Rational Condensation	Total Number of 2 (New -1) Weights*	84
Stat. Selected Subset	Total Number of Index Terms Used	71
Step 3:		
Lengthy Procedure	Total Number of Index Terms Used	89
Rational Condensation	Total Number of Index Terms Used	86
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	77
Step 4:		
Lengthy Procedure	f of TECHNICAL SKILLS	87
Rational Condensation	wf of MANAGEMENT FUNCTIONS	87
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	87
Step 5:		
Lengthy Procedure	wf of COMMUNICATION	87
Rational Condensation	wf of COMMUNICATION	91
Stat. Selected Subset	wf of PROFESSIONALISM	90
Step 6:		
Lengthy Procedure	wf of MANAGEMENT FUNCTIONS	91
Rational Condensation	wf of RECOGNITION	93
Stat. Selected Subset	wf of SKILLS AND ABILITIES	86
Step 7:		
Lengthy Procedure	f of COOPERATION	98
Rational Condensation	wf of SKILLS AND ABILITIES	95
Stat. Selected Subset	wf of REPUTE	93

* A 5 (New 3) Weight = Excellent; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 20 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED BT SAMPLES (N=164) - EVALUATION SECTION

	Variable Selected	No. of Combined BT's Classified Correctly
Step 8:		
Lengthy Procedure	wf of RESOURCEFULNESS	96
Rational Condensation	wf of CREATIVITY AND INITIATIVE	98
Stat. Selected Subset	wf of RESPONSIVENESS	98
Step 9:		
Lengthy Procedure	f of RESOURCEFULNESS	99
Rational Condensation	wf of REPRESENTATION	95
Stat. Selected Subset	wf of ASSET TO THE NAVY	99
Step 10:		
Lengthy Procedure	f of ORGANIZATION	101
Rational Condensation	wf of ORGANIZATION AND STAFFING	96
Stat. Selected Subset	wf of STAFFING	102
Step 11:		
Lengthy Procedure	f of COMMUNICATION	101
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	96
Stat. Selected Subset	wf of POTENTIAL	97
Step 12:		
Lengthy Procedure	f of PROFESSIONALISM	103
Rational Condensation	Total Number of Words in Text	94
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	98
Step 13:		
Lengthy Procedure	wf of REPUTE	103
Rational Condensation	Total Number of 1 (New -2) Weights*	93
Stat. Selected Subset	wf of DRIVE	98
Step 14:		
Lengthy Procedure	wf of PROFESSIONALISM	104
Rational Condensation	Total Number of 4 (New 2) Weights*	98
Stat. Selected Subset	wf of REPRESENTATION	98

* A 4 (New 2) Weight = Good; a 1 (New -2) Weight = Poorest.

(Continued)

TABLE 20 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED BT SAMPLES (N=164) - EVALUATION SECTION

	Variable Selected	No. of Combined BT's Classified Correctly
Step 15:		
Lengthy Procedure	f of REPRESENTATION	104
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	98
Stat. Selected Subset	wf of COMMUNICATION	94
Step 16:		
Rational Condensation	wf of ENDURANCE AND MOTIVATION	100**
Stat. Selected Subset	wf of TECHNICAL SKILLS	100
Step 17:		
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	106**
Step 54:		
Lengthy Procedure	f of PLANNING	123**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	123 out of 164 (75%)
Rational Condensation	100 out of 164 (61%)
Stat. Selected Subset	106 out of 164 (65%)

---

* A 3 (New 1) Weight = Average.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

TABLE 21

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY BT's (N=80) - JUSTIFICATION SECTION

	Variable Selected	No. of Pilot Study BT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	53
Rational Condensation	Total Number of Index Terms Used	54
Stat. Selected Subset	Total Number of Index Terms Used	50
Step 2:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	62
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	61
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	62
Step 3:		
Lengthy Procedure	f of AWARDS AND PUNISHMENT	61
Rational Condensation	wf of PLANNING-CONTROLLING	60
Stat. Selected Subset	Sum of Variables 1 through 15	59
Step 4:		
Lengthy Procedure	Sum of Variables 31 through 59	59
Rational Condensation	Total Number of 3 (New 1) Weights*	64
Stat. Selected Subset	wf of REPUTE	61
Step 5:		
Lengthy Procedure	f of REPUTE	61
Rational Condensation	Sum of Variables 1 through 15	63
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	62
Step 6:		
Lengthy Procedure	f of RESOURCEFULNESS	61
Rational Condensation	wf of CREATIVITY AND INITIATIVE	61
Stat. Selected Subset	wf of PLANNING	63
Step 7:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	61
Rational Condensation	Total Number of 2 (New -1) Weights*	62
Stat. Selected Subset	Total Number of Words in Text	63

* A 3 (New 1) Weight = Average; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 21 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP
IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION
ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES
PILOT STUDY BT's (N=80) - JUSTIFICATION SECTION

	Variable Selected	No. of Pilot Study BT's Classified Correctly
Step 8:		
Lengthy Procedure	wf of CONTROLLING	62
Rational Condensation	Total Number of Words in Text	61
Stat. Selected Subset	wf of ASSET TO THE NAVY	68
Step 9:		
Lengthy Procedure	f of SERVICE MOTIVATION	65
Rational Condensation	wf of ORGANIZATION AND STAFFING	64
Stat. Selected Subset	wf of SKILLS AND ABILITIES	66
Step 10:		
Lengthy Procedure	f of ASSET TO THE NAVY	69
Rational Condensation	Total Number of 4 (New 2) Weights*	65
Stat. Selected Subset	wf of RESPONSIVENESS	68
Step 11:		
Lengthy Procedure	f of ENDURANCE	69
Rational Condensation	wf of COMMUNICATION	65
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	69**
Step 12:		
Lengthy Procedure	Total Number of 3 (New 1) Weights*	69
Rational Condensation	wf of RECOGNITION	63
Step 13:		
Lengthy Procedure	wf of REPUTE	69
Rational Condensation	wf of SKILLS AND ABILITIES	64
Step 14:		
Lengthy Procedure	f of USE OF COMMUNICATION	69
Rational Condensation	wf of LEADERSHIP AND DIRECTING	65

* A 5 (New 3) Weight = Excellent; a 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 21 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

PILOT STUDY BT's (N=80) - JUSTIFICATION SECTION

	Variable Selected	No. of Pilot Study BT's Classified Correctly
Step 15:		
Lengthy Procedure	f of INITIATIVE	70
Rational Condensation	wf of CONDUCT AND ATTITUDE	65
Step 16:		
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	66*
Step 40:		
Lengthy Procedure	wf of RELIABILITY AND DEPENDABILITY	80*

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	80 out of 80 (100%)
Rational Condensation	66 out of 80 (83%)
Stat. Selected Subset	69 out of 80 (86%)

---

* The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

also was selected initially for all three indexing procedures in the analysis of the Justification Section for the pilot study AT's. At Step 2 in the analysis for the pilot study BT's, PRODUCTIVITY AND ACHIEVEMENT was the variable selected for all three methods, with the classification accuracy of the three methods remaining comparatively the same. From Step 3 to Step 15 the three indexing procedures diverge in the variables that were selected at each step, but certain important clusters of variables were selected by at least two of the three indexing procedures between Steps 3 and 15. These clusters were AWARDS AND PUNISHMENT/REPUTE/ASSET TO THE NAVY/RECOGNITION; PLANNING-CONTROLLING/PLANNING/CONTROLLING; sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of 3 Weights (Average); RESOURCEFULNESS/CREATIVITY AND INITIATIVE/INITIATIVE; Total Number of Words in Text; and SKILLS AND ABILITIES.

At Step 15 the lengthy indexing procedure was definitely superior to the rational condensation indexing method (88% classification accuracy compared to 81% classification accuracy), but the statistically selected subset indexing method had already attained its best classification performance at Step 11, correctly classifying 69 of the 80 pilot study BT's (86%). At Step 16 the rational condensation method reached its best classification performance, correctly classifying 66 of the 80 pilot study BT's (83%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 40 where perfect classification performance was achieved, all 80 pilot study BT's being classified correctly.

Table 22 presents the Justification Section results for the 84 cross validation BT's. As with the pilot study BT's on the Justification Section, the variable, Total Number of Index Terms Used, was selected initially for all three indexing procedures. At Step 2 LEADERSHIP AND DIRECTING was selected for all three indexing procedures, with the classification accuracy for the three methods being approximately the same. Between Steps 3 and 15 the following clusters of variables were selected for at least two of the three indexing procedures: Sum of the simple or weighted frequencies of the available set of variables for a particular indexing method; CONDUCT AND ATTITUDE/CONDUCT, INTEGRITY, AND PRIDE; PRODUCTIVITY AND ACHIEVEMENT; COOPERATION AND RESPONSIVENESS/RESPONSIVENESS; COMMUNICATION; ASSET TO THE NAVY/RECOGNITION; DRIVE; CREATIVITY AND INITIATIVE/INITIATIVE/RESOURCEFULNESS; and REPRESENTATION.

At Step 15 the lengthy indexing procedure showed an advantage over the statistically selected subset short-cut method (90% classification accuracy compared to 81% classification accuracy), but the rational condensation indexing method had already attained its best classification performance at Step 10, correctly classifying 70 of the 84 cross validation BT's (83%). At Step 19 the statistically selected subset method reached its best classification performance, correctly classifying 72 of the 84 cross validation BT's (86%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 46 where perfect classification performance was achieved, all 84 cross validation BT's being classified correctly.

A combination of the pilot study and cross validation BT samples also was reanalyzed for the Justification Section. The results of this analysis are presented in Table 23. As expected, the variable, Total Number of Index Terms Used, was selected first for all three indexing procedures just as this vari-

TABLE 22

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION BT's (N=84) - JUSTIFICATION SECTION

	Variable Selected	No. of Cross Valid. BT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	58
Rational Condensation	Total Number of Index Terms Used	58
Stat. Selected Subset	Total Number of Index Terms Used	49
Step 2:		
Lengthy Procedure	wf of LEADERSHIP AND DIRECTING	61
Rational Condensation	wf of LEADERSHIP AND DIRECTING	58
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	60
Step 3:		
Lengthy Procedure	Sum of Variables 1 through 29	58
Rational Condensation	wf of CONDUCT AND ATTITUDE	60
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	64
Step 4:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	61
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	62
Stat. Selected Subset	wf of RESPONSIVENESS	63
Step 5:		
Lengthy Procedure	wf of CONDUCT, INTEGRITY, AND PRIDE	63
Rational Condensation	wf of COMMUNICATION	64
Stat. Selected Subset	wf of ASSET TO THE NAVY	63
Step 6:		
Lengthy Procedure	f of DRIVE	66
Rational Condensation	wf of CREATIVITY AND INITIATIVE	65
Stat. Selected Subset	wf of DRIVE	64
Step 7:		
Lengthy Procedure	wf of SERVICE MOTIVATION	66
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	67
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	64

* A 3 (New 1) Weight = Average.

(Continued)

TABLE 22 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION BT's (N=84) - JUSTIFICATION SECTION

	Variable Selected	No. of Cross Valid. BT's Classified Correctly
Step 8:		
Lengthy Procedure	f of INITIATIVE	72
Rational Condensation	Sum of Variables 1 through 15	65
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	64
Step 9:		
Lengthy Procedure	f of RESPONSIVENESS	74
Rational Condensation	wf of RECOGNITION	69
Stat. Selected Subset	wf of COMMUNICATION	66
Step 10:		
Lengthy Procedure	f of ASSET TO THE NAVY	71
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	70**
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	66
Step 11:		
Lengthy Procedure	wf of DRIVE	75
Stat. Selected Subset	wf of REPRESENTATION	66
Step 12:		
Lengthy Procedure	wf of RESOURCEFULNESS	78
Stat. Selected Subset	Sum of Variables 1 through 15	66
Step 13:		
Lengthy Procedure	f of RESOURCEFULNESS	76
Stat. Selected Subset	Total Number of Words in Text	68
Step 14:		
Lengthy Procedure	f of PROFESSIONALISM	77
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	68
Step 15:		
Lengthy Procedure	f of REPRESENTATION	76
Stat. Selected Subset	wf of TECHNICAL SKILLS	68

* A 5 (New 3) Weight = Excellent; a 2 (New -1) Weight = Poor.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 22 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

CROSS VALIDATION BT's (N=84) - JUSTIFICATION SECTION

	Variable Selected	No. of Cross Valid. BT's Classified Correctly
Step 16:		
Stat. Selected Subset	wf of REPUTE	70
Step 17:		
Stat. Selected Subset	wf of SKILLS AND ABILITIES	70
Step 18:		
Stat. Selected Subset	wf of STAFFING	70
Step 19:		
Stat. Selected Subset	wf of POTENTIAL	72*
Step 46:		
Lengthy Procedure	wf of RELIABILITY AND DEPENDABILITY	84*

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	84 out of 84 (100%)
Rational Condensation	70 out of 84 (83%)
Stat. Selected Subset	72 out of 84 (86%)

---

[*] The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

TABLE 23

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED BT SAMPLES (N=164) - JUSTIFICATION SECTION

	Variable Selected	No. of Combined BT's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	112
Rational Condensation	Total Number of Index Terms Used	108
Stat. Selected Subset	Total Number of Index Terms Used	100
Step 2:		
Lengthy Procedure	wf of LEADERSHIP AND DIRECTING	112
Rational Condensation	wf of LEADERSHIP AND DIRECTING	118
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	114
Step 3:		
Lengthy Procedure	Sum of Variables 1 through 29	114
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	119
Stat. Selected Subset	wf of RESPONSIVENESS	117
Step 4:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	118
Rational Condensation	Sum of Variables 1 through 15	115
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	117
Step 5:		
Lengthy Procedure	wf of ENDURANCE	120
Rational Condensation	wf of CREATIVITY AND INITIATIVE	121
Stat. Selected Subset	Sum of Variables 1 through 15	114
Step 6		
Lengthy Procedure	f of DRIVE	116
Rational Condensation	wf of COMMUNICATION	122
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	118
Step 7:		
Lengthy Procedure	wf of AWARDS AND PUNISHMENT	121
Rational Condensation	wf of SKILLS AND ABILITIES	125
Stat. Selected Subset	Total Number of Words in Text	122

* A 3 (New 1) Weight = Average.

(Continued)

TABLE 23 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED BT SAMPLES (N=164) - JUSTIFICATION SECTION

	Variable Selected	No. of Combined BT's Classified Correctly
Step 8:		
Lengthy Procedure	wf of SKILLS AND ABILITIES	123
Rational Condensation	Total Number of 5 (New 3) Weights*	125
Stat. Selected Subset	wf of SKILLS AND ABILITIES	121
Step 9:		
Lengthy Procedure	wf of INITIATIVE	126
Rational Condensation	Total Number of 1 (New -2) Weights*	125
Stat. Selected Subset	wf of ASSET TO THE NAVY	123**
Step 10:		
Lengthy Procedure	f of RESPONSIVENESS	129
Rational Condensation	wf of ORGANIZATION AND STAFFING	126
Step 11:		
Lengthy Procedure	wf of ASSET TO THE NAVY	130
Rational Condensation	Total Number of 4 (New 2) Weights*	124
Step 12:		
Lengthy Procedure	f of CONDUCT, INTEGRITY, AND PRIDE	131
Rational Condensation	wf of RECOGNITION	126
Step 13:		
Lengthy Procedure	wf of SERVICE MOTIVATION	131
Rational Condensation	Total Number of Words in Text	128
Step 14:		
Lengthy Procedure	wf of PLANNING	133
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	129**

* A 5 (New 3) Weight = Excellent; a 4 (New 2) Weight = Good; a 1 (New -2) Weight = Poorest.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 23 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

COMBINED BT SAMPLES (N=164) - JUSTIFICATION SECTION

	Variable Selected	No. of Combined BT's Classified Correctly
Step 66:		
Lengthy Procedure	f of POTENTIAL	150*

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	150 out of 164 (91%)
Rational Condensation	129 out of 164 (79%)
Stat. Selected Subset	123 out of 164 (75%)

---

* The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

able universally was selected first when the two BT samples were analyzed separately. The variable, Total Number of Index Terms Used, also was the first variable selected for the three AT analyses on the Justification Section. This variable reflects the variety of specific areas of an individual's performance that the evaluator chose to comment on, and is measured by the number of different index terms chosen by the indexer to encompass the narrative content of the Evaluation or Justification Section. It appears that the range of skills and abilities that a chief petty officer possesses is a key factor in his superior performance.

At Step 2 in Table 23, LEADERSHIP AND DIRECTING was selected for all three indexing procedures. Between Steps 3 and 15 the following clusters of variables were selected for at least two of the three indexing procedures: Sum of the simple or weighted frequencies of the available set of variables for a particular indexing method; PRODUCTIVITY AND ACHIEVEMENT; RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; CREATIVITY AND INITIATIVE/INITIATIVE; AWARDS AND PUNISHMENT/ASSET TO THE NAVY/RECOGNITION; SKILLS AND ABILITIES; and Total Number of Words in Text.

The three indexing procedures exhibit similar classification accuracy through Step 9 at which point the statistically selected subset method achieved its best classification performance, correctly classifying 123 of the 164 combined BT's (75%). At Step 14 the rational condensation indexing method attained its best classification performance, correctly classifying 129 of the 164 combined BT's (79%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 66 where it finally achieved a classification accuracy of 91%, correctly classifying 150 of the 164 combined BT's.

CS Comparison

CS Comparison - Evaluation Section. Table 24 shows a comparison of the performance of the three indexing procedures for the 60 generalization CS's on the Evaluation Section of the Evaluation Report. At Step 1 the similar variables, CONTROLLING and PLANNING-CONTROLLING, were selected for the lengthy indexing procedure and the rational condensation indexing method, respectively. ASSET TO THE NAVY was selected for the statistically selected subset method at Step 1 and for the lengthy indexing procedure at Step 2. Within Steps 2 and 3 the variable, MANAGEMENT FUNCTIONS, was selected for all three indexing procedures. Between Steps 3 and 15 the following clusters of variables were selected for at least two of the three indexing procedures: Total Number of 3 Weights (Average); SKILLS AND ABILITIES; sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of Index Terms Used; PLANNING; ENDURANCE/SERVICE MOTIVATION/DRIVE/ENDURANCE AND MOTIVATION; CREATIVITY AND INITIATIVE/INITIATIVE; ORGANIZATION AND STAFFING/ORGANIZATION; POTENTIAL; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS; LEADERSHIP AND DIRECTING; and REPRESENTATION.

The three indexing procedures exhibit similar classification accuracy through Step 15 at which point the rational condensation indexing method achieved its best classification performance, correctly classifying 48 of the 60 generalization CS's (80%). At Step 16 the statistically selected subset method reached its best classification performance, correctly classifying 46

TABLE 24

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION CS's (N=60) - EVALUATION SECTION

	Variable Selected	No. of Generalizat'n CS's Classified Correctly
Step 1:		
Lengthy Procedure	wf of CONTROLLING	20
Rational Condensation	wf of PLANNING-CONTROLLING	23
Stat. Selected Subset	wf of ASSET TO THE NAVY	20
Step 2:		
Lengthy Procedure	wf of ASSET TO THE NAVY	28
Rational Condensation	wf of MANAGEMENT FUNCTIONS	25
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	29
Step 3:		
Lengthy Procedure	f of MANAGEMENT FUNCTIONS	30
Rational Condensation	Total Number of Words in Text	33
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	35
Step 4:		
Lengthy Procedure	wf of SKILLS AND ABILITIES	30
Rational Condensation	Sum of Variables 1 through 15	37
Stat. Selected Subset	wf of SKILLS AND ABILITIES	35
Step 5:		
Lengthy Procedure	Total Number of Index Terms Used	32
Rational Condensation	Total Number of Index Terms Used	37
Stat. Selected Subset	wf of PLANNING	38
Step 6:		
Lengthy Procedure	f of ENDURANCE	34
Rational Condensation	wf of CREATIVITY AND INITIATIVE	39
Stat. Selected Subset	wf of REPUTE	41
Step 7:		
Lengthy Procedure	Sum of Variables 31 through 59	38
Rational Condensation	wf of ORGANIZATION AND STAFFING	41
Stat. Selected Subset	Total Number of Index Terms Used	38

* A 3 (New 1) Weight = Average.

(Continued)

TABLE 24 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION CS's (N=60) - EVALUATION SECTION

	Variable Selected	No. of Generalizat'n CS's Classified Correctly
Step 8:		
Lengthy Procedure	wf of ORGANIZATION	39
Rational Condensation	Total Number of 3 (New 1) Weights*	39
Stat. Selected Subset	Sum of Variables 1 through 15	39
Step 9:		
Lengthy Procedure	wf of POTENTIAL	42
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	40
Stat. Selected Subset	wf of TECHNICAL SKILLS	45
Step 10:		
Lengthy Procedure	wf of PLANNING	43
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	44
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	43
Step 11:		
Lengthy Procedure	f of SERVICE MOTIVATION	46
Rational Condensation	wf of CONDUCT AND ATTITUDE	45
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	44
Step 12:		
Lengthy Procedure	f of PLANNING	41
Rational Condensation	Total Number of 4 (New 2) Weights*	45
Stat. Selected Subset	wf of POTENTIAL	41
Step 13:		
Lengthy Procedure	Total Number of 3 (New 1) Weights*	44
Rational Condensation	wf of COMMUNICATION	47
Stat. Selected Subset	wf of DRIVE	44
Step 14:		
Lengthy Procedure	f of LEADERSHIP AND DIRECTING	43
Rational Condensation	wf of REPRESENTATION	47
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	44

* A 5 (New 3) Weight = Excellent; a 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average.

(Continued)

TABLE 24 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION CS's (N=60) - EVALUATION SECTION

	Variable Selected	No. of Generalizat'n CS's Classified Correctly
Step 15:		
Lengthy Procedure	f of INITIATIVE	44
Rational Condensation	wf of ENDURANCE AND MOTIVATION	48*
Stat. Selected Subset	wf of REPRESENTATION	44
Step 16:		
Stat. Selected Subset	wf of RESPONSIVENESS	46*
Step 44:		
Lengthy Procedure	f of COMMUNICATION	58*

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	58 out of 60 (97%)
Rational Condensation	48 out of 60 (80%)
Stat. Selected Subset	46 out of 60 (77%)

---

* The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

of the 60 generalization CS's (77%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 44 where it finally attained a classification accuracy of 97%, correctly classifying 58 of the 60 generalization CS's.

CS Comparison - Justification Section. Table 25 presents a comparison of the performance of the three indexing procedures for the 60 generalization CS's on the Justification Section of the Evaluation Report. For all three indexing procedures, the first variable selected was Total Number of Index Terms Used, with the classification accuracy for the three methods being approximately the same initially. At Step 2 the similar variables, PROFESSIONALISM and PROFESSIONAL AND TECHNICAL SKILLS, were selected for all three indexing procedures. Between Steps 3 and 13 the following clusters of variables were selected by at least two of the three indexing procedures: INITIATIVE/CREATIVITY AND INITIATIVE; COMMUNICATION; COOPERATION AND RESPONSIVENESS/COOPERATION; REPRESENTATION; STAFFING/ORGANIZATION AND STAFFING; POTENTIAL; Total Number of Words in Text; PRODUCTIVITY AND ACHIEVEMENT; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; and PLANNING.

The three indexing procedures exhibit similar classification accuracy through Step 11 at which point the statistically selected subset indexing method achieved its best classification performance, correctly classifying 52 of the 60 generalization CS's (87%). At Step 13 the rational condensation indexing method reached its best classification performance, correctly classifying 55 of the 60 generalization CS's (92%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 35 where perfect classification performance was achieved, all 60 generalization CS's being classified correctly.

RM Comparison

RM Comparison - Evaluation Section. Table 26 shows a comparison of the performance of the three indexing procedures for the 162 generalization RM's on the Evaluation Section of the Evaluation Report. At Step 1 the variable, Total Number of 2 Weights (Poor), was selected for all three indexing procedures, with the classification accuracy for the three methods being approximately the same initially. Between Steps 2 and 15 a number of important clusters of variables were selected for at least two of the three indexing procedures. These clusters were AWARDS AND PUNISHMENT/RECOGNITION/POTENTIAL/REPUTE/ASSET TO THE NAVY; MANAGEMENT FUNCTIONS; Total Number of 5 Weights (Excellent); COMMUNICATION; RELIABILITY AND DEPENDABILITY/CONDUCT AND ATTITUDE; INTELLECTUAL FUNCTIONING; Total Number of Index Terms Used; PRODUCTIVITY AND ACHIEVEMENT; REPRESENTATION; and RESPONSIVENESS.

At Step 15 the lengthy indexing procedure was slightly superior to the two short-cut indexing methods (65% classification accuracy for the lengthy procedure compared to 60% classification accuracy for both the rational condensation method and the statistically selected subset method). At Step 17 the rational condensation indexing method achieved its best classification performance, correctly classifying 102 of the 162 generalization RM's (63%). At Step 20 the statistically selected subset indexing method reached its best classification performance, correctly classifying 105 of the 162 generalization

TABLE 25

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION CS's (N=60) - JUSTIFICATION SECTION

	Variable Selected	No. of Generalizat'n CS's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	42
Rational Condensation	Total Number of Index Terms Used	44
Stat. Selected Subset	Total Number of Index Terms Used	45
Step 2:		
Lengthy Procedure	f of PROFESSIONALISM	44
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	46
Stat. Selected Subset	wf of PROFESSIONALISM	44
Step 3:		
Lengthy Procedure	f of INITIATIVE	44
Rational Condensation	wf of CREATIVITY AND INITIATIVE	49
Stat. Selected Subset	wf of COMMUNICATION	45
Step 4:		
Lengthy Procedure	wf of COMMUNICATION	47
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	50
Stat. Selected Subset	wf of REPRESENTATION	46
Step 5:		
Lengthy Procedure	wf of REPRESENTATION	47
Rational Condensation	wf of REPRESENTATION	50
Stat. Selected Subset	wf of STAFFING	47
Step 6:		
Lengthy Procedure	wf of COOPERATION	47
Rational Condensation	wf of CONDUCT AND ATTITUDE	50
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	49
Step 7:		
Lengthy Procedure	f of CONTROLLING	50
Rational Condensation	wf of LEADERSHIP AND DIRECTING	52
Stat. Selected Subset	wf of POTENTIAL	49

---

* A 2 (New -1) Weight = Poor.

(Continued)

TABLE 25 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION CS's (N=60) - JUSTIFICATION SECTION

	Variable Selected	No. of Generalizat'n CS's Classified Correctly
Step 8:		
Lengthy Procedure	f of POTENTIAL	51
Rational Condensation	wf of COMMUNICATION	51
Stat. Selected Subset	Total Number of Words in Text	51
Step 9:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	55
Rational Condensation	Total Number of 3 (New 1) Weights*	51
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	51
Step 10:		
Lengthy Procedure	wf of SKILLS AND ABILITIES	53
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	51
Stat. Selected Subset	wf of PLANNING	50
Step 11:		
Lengthy Procedure	wf of PLANNING	54
Rational Condensation	Total Number of Words in Text	53
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	52**
Step 12:		
Lengthy Procedure	Total Number of Words in Text	54
Rational Condensation	wf of SKILLS AND ABILITIES	54
Step 13:		
Lengthy Procedure	wf of PROFESSIONALISM	55
Rational Condensation	wf of ORGANIZATION AND STAFFING	55**
Step 35:		
Lengthy Procedure	wf of POTENTIAL	60**

---

* A 3 (New 1) Weight = Average.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 25 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION CS's (N=60) - JUSTIFICATION SECTION

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	60 out of 60 (100%)
Rational Condensation	55 out of 60 (92%)
Stat. Selected Subset	52 out of 60 (87%)

TABLE 26

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION RM's (N=162) - EVALUATION SECTION

	Variable Selected	No. of Generalizat'n RM's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of 2 (New -1) Weights*	70
Rational Condensation	Total Number of 2 (New -1) Weights*	70
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	69
Step 2:		
Lengthy Procedure	wf of AWARDS AND PUNISHMENT	79
Rational Condensation	wf of RECOGNITION	85
Stat. Selected Subset	wf of POTENTIAL	78
Step 3:		
Lengthy Procedure	wf of POTENTIAL	87
Rational Condensation	wf of MANAGEMENT FUNCTIONS	83
Stat. Selected Subset	Total Number of 5 (New 3) Weights*	85
Step 4:		
Lengthy Procedure	wf of MANAGEMENT FUNCTIONS	88
Rational Condensation	wf of COMMUNICATION	80
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	87
Step 5:		
Lengthy Procedure	Total Number of 5 (New 3) Weights*	91
Rational Condensation	Total Number of 5 (New 3) Weights*	82
Stat. Selected Subset	wf of COMMUNICATION	87
Step 6:		
Lengthy Procedure	f of RELIABILITY AND DEPENDABILITY	100
Rational Condensation	wf of CONDUCT AND ATTITUDE	85
Stat. Selected Subset	wf of REPUTE	90
Step 7:		
Lengthy Procedure	f of COMMUNICATION	103
Rational Condensation	wf of INTELLECTUAL FUNCTIONING	82
Stat. Selected Subset	Total Number of Index Terms Used	93

---

* A 5 (New 3) Weight = Excellent; a 2 (New -1) Weight = Poor.

(Continued)

TABLE 26 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION RM's (N=162) - EVALUATION SECTION

	Variable Selected	No. of Generalizat'n RM's Classified Correctly
Step 8:		
Lengthy Procedure	f of POTENTIAL	99
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	89
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	95
Step 9:		
Lengthy Procedure	wf of REPUTE	95
Rational Condensation	Total Number of Index Terms Used	91
Stat. Selected Subset	wf of PROFESSIONALISM	96
Step 10:		
Lengthy Procedure	f of REPUTE	94
Rational Condensation	Sum of Variables 1 through 15	95
Stat. Selected Subset	wf of REPRESENTATION	98
Step 11:		
Lengthy Procedure	f of INTELLECTUAL FUNCTIONING	92
Rational Condensation	wf of REPRESENTATION	94
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	95
Step 12:		
Lengthy Procedure	f of RESPONSIVENESS	98
Rational Condensation	wf of PLANNING-CONTROLLING	97
Stat. Selected Subset	wf of ASSET TO THE NAVY	94
Step 13:		
Lengthy Procedure	wf of RESPONSIVENESS	99
Rational Condensation	wf of LEADERSHIP AND DIRECTING	97
Stat. Selected Subset	wf of STAFFING	101
Step 14:		
Lengthy Procedure	wf of SERVICE MOTIVATION	99
Rational Condensation	Total Number of Words in Text	99
Stat. Selected Subset	wf of RESPONSIVENESS	101

[*] A 3 (New 1) Weight = Average.

(Continued)

TABLE 26 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION RM's (N=162) - EVALUATION SECTION

	Variable Selected	No. of Generalizat'n RM's Classified Correctly
Step 15:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	105
Rational Condensation	wf of SKILLS AND ABILITIES	97
Stat. Selected Subset	wf of DRIVE	98
Step 16:		
Rational Condensation	Total Number of 4 (New 2) Weights*	98
Stat. Selected Subset	Total Number of Words in Text	100
Step 17:		
Rational Condensation	wf of ORGANIZATION AND STAFFING	102**
Stat. Selected Subset	wf of LEADERSHIP AND DIRECTING	100
Step 18:		
Stat. Selected Subset	wf of SKILLS AND ABILITIES	103
Step 19:		
Stat. Selected Subset	wf of PLANNING	103
Step 20:		
Lengthy Procedure	wf of PLANNING	107
Stat. Selected Subset	Sum of Variables 1 through 15	105**
Step 63:		
Lengthy Procedure	f of INTELLECTUAL FUNCTIONING *(Variable Removed)*	131**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	131 out of 162 (81%)
Rational Condensation	102 out of 162 (63%)
Stat. Selected Subset	105 out of 162 (65%)

* A 4 (New 2) Weight = Good.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

RM's (65%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 63 where it finally attained a classification accuracy of 81%, correctly classifying 131 of the 162 generalization RM's.

RM Comparison - Justification Section. Table 27 presents a comparison of the performance of the three indexing procedures for the 162 generalization RM's on the Justification Section of the Evaluation Report. For all three indexing procedures, the first variable selected was Total Number of Index Terms Used. Between Steps 2 and 15 the following important clusters of variables were selected for at least two of the three indexing procedures: Sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of 4 Weights (Good); PRODUCTIVITY AND ACHIEVEMENT; DRIVE/ENDURANCE AND MOTIVATION/ENDURANCE; COOPERATION/RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; CONDUCT AND ATTITUDE/GROOMING AND ATTIRE; STAFFING/ORGANIZATION AND STAFFING; Total Number of 3 Weights (Average); REPUTE; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS/PROFESSIONALISM; PLANNING; CREATIVITY AND INITIATIVE/INITIATIVE; and POTENTIAL.

The rational condensation indexing method achieved its best classification performance at Step 11, correctly classifying 130 of the 162 generalization RM's (80%). At Step 16 the statistically selected subset method reached its best classification performance, correctly classifying 129 of the 162 generalization RM's (80%). The stepwise discriminant analysis for the lengthy indexing procedure continued on to Step 40 where it finally achieved a classification accuracy of 89%, correctly classifying 144 of the 162 generalization RM's.

Summary and Conclusions

The two short-cut indexing procedures that were developed for this study compared favorably with the classification accuracy achieved by the original lengthy indexing procedure in the early steps of the stepwise discriminant analysis process, i.e., between Steps 1 and 10. Beyond Step 10 the lengthy indexing procedure, with its greater complement of available variables, typically displayed a superior classification performance as the stepwise discriminant analysis process continued to try to maximize its classification accuracy. In all of the comparisons that were made, the lengthy indexing procedure exceeded the better classification performance of the two short-cut indexing methods. However, since the lengthy indexing procedure provided more variables to the stepwise discriminant analysis process, it was expected that this method would demonstrate superior classification performance. There is other evidence that most of the discrimination which is achievable can be attributed to the variables selected early by the stepwise discriminant anaysis process. In previous research with the lengthy indexing procedure, when the discriminant functions developed on one sample were used to classify a second cross validation sample, the classification performance of the lengthy procedure dropped markedly, typically from near perfect classification for the original sample to 65-70% classification accuracy for the cross validation sample.[13] This level of cross validation classification accuracy was achieved early in the stepwise discriminant analysis process, typically by the fifth step. This important finding from a previous study indicated that the variables selected by the stepwise discriminant analysis program for the early

TABLE 27

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION RM's (N=162) - JUSTIFICATION SECTION

	Variable Selected	No. of Generalizat'n RM's Classified Correctly
Step 1:		
Lengthy Procedure	Total Number of Index Terms Used	103
Rational Condensation	Total Number of Index Terms Used	100
Stat. Selected Subset	Total Number of Index Terms Used	112
Step 2:		
Lengthy Procedure	Sum of Variables 31 through 59	105
Rational Condensation	Total Number of 4 (New 2) Weights*	103
Stat. Selected Subset	Sum of Variables 1 through 15	105
Step 3:		
Lengthy Procedure	wf of PRODUCTIVITY AND ACHIEVEMENT	110
Rational Condensation	wf of PRODUCTIVITY AND ACHIEVEMENT	109
Stat. Selected Subset	wf of PRODUCTIVITY AND ACHIEVEMENT	113
Step 4:		
Lengthy Procedure	f of PRODUCTIVITY AND ACHIEVEMENT	114
Rational Condensation	Total Number of 5 (New 3) Weights*	112
Stat. Selected Subset	wf of DRIVE	116
Step 5:		
Lengthy Procedure	f of COOPERATION	117
Rational Condensation	wf of CONDUCT AND ATTITUDE	117
Stat. Selected Subset	wf of RESPONSIVENESS	114
Step 6:		
Lengthy Procedure	wf of GROOMING AND ATTIRE	123
Rational Condensation	wf of ENDURANCE AND MOTIVATION	121
Stat. Selected Subset	wf of STAFFING	113
Step 7:		
Lengthy Procedure	wf of STAFFING	127
Rational Condensation	wf of COOPERATION AND RESPONSIVENESS	123
Stat. Selected Subset	Total Number of 3 (New 1) Weights*	117

* A 5 (New 3) Weight = Excellent; a 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average.

(Continued)

TABLE 27 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION RM's (N=162) - JUSTIFICATION SECTION

	Variable Selected	No. of Generalizat'n RM's Classified Correctly
Step 8:		
Lengthy Procedure	wf of ENDURANCE	130
Rational Condensation	Total Number of 3 (New 1) Weights*	123
Stat. Selected Subset	wf of REPUTE	118
Step 9:		
Lengthy Procedure	f of STAFFING	131
Rational Condensation	wf of PROFESSIONAL AND TECHNICAL SKILLS	126
Stat. Selected Subset	wf of PLANNING	123
Step 10:		
Lengthy Procedure	wf of PLANNING	134
Rational Condensation	wf of CREATIVITY AND INITIATIVE	129
Stat. Selected Subset	Total Number of Words in Text	125
Step 11:		
Lengthy Procedure	Total Number of 4 (New 2) Weights*	129
Rational Condensation	wf of ORGANIZATION AND STAFFING	130**
Stat. Selected Subset	wf of MANAGEMENT FUNCTIONS	125
Step 12:		
Lengthy Procedure	f of POTENTIAL	130
Stat. Selected Subset	wf of SKILLS AND ABILITIES	125
Step 13:		
Lengthy Procedure	f of INITIATIVE	130
Stat. Selected Subset	wf of ASSET TO THE NAVY	125
Step 14:		
Lengthy Procedure	wf of TECHNICAL SKILLS	133
Stat. Selected Subset	wf of POTENTIAL	128

* A 4 (New 2) Weight = Good; a 3 (New 1) Weight = Average.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

(Continued)

TABLE 27 (CONT.)

COMPARISON OF THE VARIABLES SELECTED AT EACH STEP IN THE STEPWISE DISCRIMINANT ANALYSIS AND THE CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES

GENERALIZATION RM's (N=162) - JUSTIFICATION SECTION

	Variable Selected	No. of Generalizat'n RM's Classified Correctly
Step 15:		
Lengthy Procedure	f of REPUTE	132
Stat. Selected Subset	wf of PROFESSIONALISM	128
Step 16:		
Stat. Selected Subset	Total Number of 2 (New -1) Weights*	129**
Step 40:		
Lengthy Procedure	wf of FLEXIBILITY	144**

RECAPITULATION OF CLASSIFICATION ACCURACY FOR THE THREE INDEXING PROCEDURES:

Lengthy Procedure	144 out of 162 (89%)
Rational Condensation	130 out of 162 (80%)
Stat. Selected Subset	129 out of 162 (80%)

---

* A 2 (New -1) Weight = Poor.

** The *underscore* indicates the best classification achieved in the stepwise discriminant analysis for a particular indexing procedure.

steps in the discriminant analysis are crucial variables, playing a major role in differentiating among the three criterion groups.

Perhaps, then, a more meaningful comparison among the three indexing procedures is the classification performance that they achieved between Steps 10 and 20 in the stepwise discriminant analysis process, the range of steps at which the two short-cut methods attained their best classification accuracy. In all of the comparisons reported in this section, the classification performance of the three indexing procedures was similar between Steps 10 and 20, with the lengthy procedure typically having a slight but definite edge over the two short-cut methods. In some comparisons the rational condensation indexing method, at its best classification accuracy, demonstrated superior classification performance to the best performance of the statistically selected subset indexing method, but in other comparisons the statistically selected subset method performed better (see Table 28). In eight of the 16 comparisons that were made, the rational condensation method achieved better classification accuracy. In seven of the 16 comparisons, the statistically selected subset method attained better classification accuracy. In one comparison the two short-cut indexing methods performed equally well. Therefore, the criterion that was adopted to determine which of the two short-cut methods should be considered superior and elected as the preferred method for subsequent research studies was how well each short-cut method tracked the original lengthy indexing procedure in selecting variables into the discriminant function. Of the two short-cut indexing methods, the one that from the initial step more faithfully tracked the original lengthy indexing procedure in selecting variables into the discriminant function was the rational condensation method. Moreover, the rational condensation method examines *all* of the information contained in a narrative performance evaluation in contrast to the statistically selected subset method which takes into consideration only portions of the narrative text. In the rational condensation short-cut method the indexer is obligated to make an indexing decision for each segment of narrative text. The logical choice for any one segment of text usually is between only two of the 15 index terms comprising this short-cut method and between only two of the five possible weights, thus increasing the likelihood that a correct indexing decision will be made. However, the statistically selected subset method requires that the indexer ignore portions of the text whose content does not map onto one of the 15 index terms available for this short-cut method, 14 of the 29 original index terms having been eliminated from the statistically selected subset. If none of the 15 available terms appears to fit a segment of narrative text, then this particular segment is left unindexed. Therefore, a 2-step indexing decision is required for the statistically selected subset method: (1) whether or not to index a particular segment of narrative text, and (2) if the decision is made to index, then the choice of an appropriate index term from the 15 available terms and the choice of a weight from the five possibilities. The decision *not* to index a particular segment of narrative text may be an easier choice than the decision of which index term and weight to use. Consequently, it was felt that the statistically selected subset method is more subject to indexing error and inconsistency than the rational condensation method. Because of this hazard with using the statistically selected subset method and because the rational condensation method more faithfully tracked the original lengthy indexing procedure in selecting variables into the discriminant function, the rational condensation method was chosen as the preferred short-cut indexing method for further research investigations.

TABLE 28

SUMMARY OF THE BEST CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES*

		Classification Accuracy				
			N Correct at Best Step			
Occupational Specialty-Section/Sample	Indexing Procedure	N Correct at Step 1	Step No.	N Correct	% Correct	Source Table
AT-Evaluation						
Pilot Study N=144	Lengthy Procedure	68	50	105	73%	12
	Rational Condensation	68	18	80	56%	
	Stat. Selected Subset	67	14	83	58%	
Cross Valid. N=138	Lengthy Procedure	58	48	110	80%	13
	Rational Condensation	61	18	96	70%	
	Stat. Selected Subset	59	16	87	63%	
Combined N=282	Lengthy Procedure	129	46	184	65%	14
	Rational Condensation	129	6	166	59%	
	Stat. Selected Subset	130	21	159	56%	
AT-Justification						
Pilot Study N-144	Lengthy Procedure	95	58	137	95%	15
	Rational Condensation	96	18	110	76%	
	Stat. Selected Subset	99	15	116	81%	
Cross Valid. N=138	Lengthy Procedure	88	46	129	93%	16
	Rational Condensation	88	16	110	80%	
	Stat. Selected Subset	85	18	109	79%	
Combined N=282	Lengthy Procedure	184	46	230	82%	17
	Rational Condensation	182	17	217	77%	
	Stat. Selected Subset	181	12	214	76%	

* The best classification accuracy for the better of the two short-cut indexing procedures in each of the 16 comparisons that was made is underscored. In eight of the 16 comparisons, the rational condensation method achieved better classification accuracy. In seven of the 16 comparisons, the statistically selected subset method attained better classification accuracy. In one comparison the two short-cut indexing methods performed equally well.

(Continued)

TABLE 28 (CONT.)

SUMMARY OF THE BEST CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES*

Occupational Specialty-Section/Sample	Indexing Procedure	Classification Accuracy				Source Table
			N Correct at Best Step			
		N Correct at Step 1	Step No.	N Correct	% Correct	
BT-Evaluation						
Pilot Study N=80	Lengthy Procedure	34	50	77	96%	18
	Rational Condensation	34	16	57	71%	
	Stat. Selected Subset	32	12	57	71%	
Cross Valid. N=84	Lengthy Procedure	39	62	82	98%	19
	Rational Condensation	39	14	63	75%	
	Stat. Selected Subset	33	21	64	76%	
Combined N=164	Lengthy Procedure	73	54	123	75%	20
	Rational Condensation	73	16	100	61%	
	Stat. Selected Subset	65	17	106	65%	
BT- Justification						
Pilot Study N=80	Lengthy Procedure	53	40	80	100%	21
	Rational Condensation	54	16	66	83%	
	Stat. Selected Subset	50	11	69	86%	
Cross Valid. N=84	Lengthy Procedure	58	46	84	100%	22
	Rational Condensation	58	10	70	83%	
	Stat. Selected Subset	49	19	72	86%	
Combined N=164	Lengthy Procedure	112	66	150	91%	23
	Rational Condensation	108	14	129	79%	
	Stat. Selected Subset	100	9	123	75%	

* The best classification accuracy for the better of the two short-cut indexing procedures in each of the 16 comparisons that was made is underscored. In eight of the 16 comparisons, the rational condensation method achieved better classification accuracy. In seven of the 16 comparisons, the statistically selected subset method attained better classification accuracy. In one comparison the two short-cut indexing methods performed equally well.

(Continued)

TABLE 28 (CONT.)

SUMMARY OF THE BEST CLASSIFICATION ACCURACY ACHIEVED BY THE THREE INDEXING PROCEDURES*

Occupational Specialty-Section/Sample	Indexing Procedure	Classification Accuracy				Source Table
			N Correct at Best Step			
		N Correct at Step 1	Step No.	N Correct	% Correct	
CS-Evaluation						
Generalization N=60	Lengthy Procedure	20	44	58	97%	24
	Rational Condensation	23	15	48	80%	
	Stat. Selected Subset	20	16	46	77%	
CS-Justification						
Generalization N=60	Lengthy Procedure	42	35	60	100%	25
	Rational Condensation	44	13	55	92%	
	Stat. Selected Subset	45	11	52	87%	
RM-Evaluation						
Generalization N=162	Lengthy Procedure	70	63	131	81%	26
	Rational Condensation	70	17	102	63%	
	Stat. Selected Subset	69	20	105	65%	
RM-Justification						
Generalization N=162	Lengthy Procedure	103	40	144	89%	27
	Rational Condensation	100	11	130	80%	
	Stat. Selected Subset	112	16	129	80%	

---

* The best classification accuracy for the better of the two short-cut indexing procedures in each of the 16 comparisons that was made is underscored. In eight of the 16 comparisons, the rational condensation method achieved better classification accuracy. In seven of the 16 comparisons, the statistically selected subset method attained better classification accuracy. In one comparison the two short-cut indexing methods performed equally well.

When the number of predictor variables is large in relation to the number of cases (the worst instance in this study being 67 variables for the lengthy indexing procedure as applied to the 60 generalization CS's), the solution achieved by the stepwise discriminant analysis algorithm, as in the case of multiple regression, may converge on a set of predictor variables that solves the classification problem perfectly for that particular sample, but may not constitute the same set of variables that might be selected for another sample or for another indexing procedure, a different set of variables also being able to achieve perfect or near perfect classification.[14] Therefore, it is extremely interesting to note that for all three indexing procedures, the key variables selected early in the stepwise discriminant analysis process for the Evaluation Section were Total Number of 5 Weights (Excellent) and Total Number of 2 Weights (Poor). This was true for all comparisons made on the Evaluation Section of the Evaluation Report except for the 60 generalization CS's. This finding points up the need to cross validate the results of studies based on small N's where the number of predictor variables exceeds the number of cases in order to determine which discriminating variables are constant over several samples. The conclusion that can be drawn from these findings is that the modifying adjectives and adverbs used by an evaluator to assess an individual's performance in the Evaluation Section of the Evaluation Report are key factors in distinguishing between superior performance and less stellar achievements, regardless of the occupational specialty being analyzed, with the exception of the 60-case generalization CS sample which constituted the worst case statistically for finding a valid, reproducible set of predictor variables.

When one examines the results for the Justification Section of the Evaluation Report, the findings are unequivocal. *Without exception* for all comparisons made, the first variable selected for the Justification Section was Total Number of Index Terms Used. This variable reflects the variety of specific areas of an individual's performance that the evaluator chose to comment on, and is measured by the number of different index terms chosen by the indexer to encompass the narrative content. This finding indicates that the range of skills and abilities that a chief petty officer manifests is a key factor in his superior performance as narrated by the evaluator in the Justification Section. Another finding, which corroborates the results of previous research[15], is that *without exception* better classification was achieved in the content analysis of the narrative comments in the Justification Section compared to the Evaluation Section, regardless of which of the three indexing procedures was employed.

The results from an earlier research study indicated that classification procedures based on the lengthy content analysis methodology should be tailored to specific occupations.[16] The findings from the study being reported here substantiate the earlier research results and show that for each occupational specialty on a particular section of the Evaluation Report, the variables selected for at least two of the three indexing procedures were identical and not necessarily the same as those variables selected for a different occupational specialty. A summary enumeration of these key discriminating variables selected in the first 15 steps by the stepwise discriminant analysis procedure for each occupational specialty is given below.

AT's - Key Discriminating Variables for the Evaluation Section. The following list of variables was determined from the results of the combined

AT analysis on the Evaluation Section (N=282). The key discriminating clusters of variables were Total Number of 5 Weights (Excellent); Total Number of 2 Weights (Poor); LEADERSHIP AND DIRECTING; TECHNICAL SKILLS/PROFESSIONAL AND TECHNICAL SKILLS; MANAGEMENT FUNCTIONS; Total Number of 3 Weights (Average): RESPONSIVENESS; COMMUNICATION; POTENTIAL; and DRIVE.

AT's - Key Discriminating Variables for the Justification Section. The following list of variables was determined from the results of the combined AT analysis on the Justification Section (N=282). The key discriminating clusters of variables were Total Number of Index Terms Used; sum of the weighted frequencies of the available set of variables for a particular indexing procedure; PRODUCTIVITY AND ACHIEVEMENT; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS/PROFESSIONALISM; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; STAFFING/ORGANIZATION AND STAFFING; INTELLECTUAL FUNCTIONING; ENDURANCE AND MOTIVATION/DRIVE; MANAGEMENT FUNCTIONS; REPUTE; Total Number of 2 Weights (Poor); COMMUNICATION; REPRESENTATION; and Total Number of Words in Text.

BT's - Key Discriminating Variables for the Evaluation Section. The following list of variables was determined from the results of the combined BT analysis on the Evaluation Section (N=164). The key discriminating clusters of variables were Total Number of 5 Weights (Excellent); Total Number of 2 Weights (Poor); Total Number of Index Terms Used; MANAGEMENT FUNCTIONS; COMMUNICATION; PROFESSIONALISM; RECOGNITION/REPUTE/ASSET TO THE NAVY/POTENTIAL; SKILLS AND ABILITIES; RESOURCEFULNESS/CREATIVITY AND INITIATIVE; REPRESENTATION; ORGANIZATION/STAFFING/ORGANIZATION AND STAFFING; and PRODUCTIVITY AND ACHIEVEMENT.

BT's - Key Discriminating Variables for the Justification Section. The following list of variables was determined from the results of the combined BT analysis on the Justification Section (N=164). The key discriminating clusters of variables were Total Number of Index Terms Used; LEADERSHIP AND DIRECTING; sum of the simple or weighted frequencies of the available set of variables for a particular indexing method; PRODUCTIVITY AND ACHIEVEMENT; RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; CREATIVITY AND INITIATIVE/INITIATIVE; AWARDS AND PUNISHMENT/ASSET TO THE NAVY/RECOGNITION; SKILLS AND ABILITIES; and Total Number of Words in Text.

CS's - Key Discriminating Variables for the Evaluation Section. The key discriminating clusters of variables for the 60 generalization CS's on the Evaluation Section were CONTROLLING/PLANNING-CONTROLLING; ASSET TO THE NAVY; MANAGEMENT FUNCTIONS; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of Index Terms Used; PLANNING; ENDURANCE/SERVICE MOTIVATION/DRIVE/ENDURANCE AND MOTIVATION; CREATIVITY AND INITIATIVE/INITIATIVE; ORGANIZATION AND STAFFING/ORGANIZATION; POTENTIAL; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS; LEADERSHIP AND DIRECTING; and REPRESENTATION.

CS's - Key Discriminating Variables for the Justification Section. The key discriminating clusters of variables for the 60 generalization CS's on the Justification Section were Total Number of Index Terms Used; PROFESSIONALISM/PROFESSIONAL AND TECHNICAL SKILLS; INITIATIVE/CREATIVITY AND INITIATIVE; COMMUNICATION; COOPERATION AND RESPONSIVENESS/COOPERATION; REPRESENTATION;

STAFFING/ORFANIZATION AND STAFFING; POTENTIAL; Total Number of Words in Text; PRODUCTIVITY AND ACHIEVEMENT; Total Number of 3 Weights (Average); SKILLS AND ABILITIES; and PLANNING.

RM's - Key Discriminating Variables for the Evaluation Section. The key discriminating clusters of variables for the 162 generalization RM's on the Evaluation Section were Total Number of 2 Weights (Poor); AWARDS AND PUNISHMENT/RECOGNITION/POTENTIAL/REPUTE/ASSET TO THE NAVY; MANAGEMENT FUNCTIONS; Total Number of 5 Weights (Excellent); COMMUNICATION; RELIABILITY AND DEPENDABILITY/CONDUCT AND ATTITUDE; INTELLECTUAL FUNCTIONING; Total Number of Index Terms Used; PRODUCTIVITY AND ACHIEVEMENT; REPRESENTATION; and RESPONSIVENESS.

RM's - Key Discriminating Variables for the Justification Section. The key discriminating clusters of variables for the 162 generalization RM's on the Justification Section were Total Number of Index Terms Used; sum of the weighted frequencies of the available set of variables for a particular indexing method; Total Number of 4 Weights (Good); PRODUCTIVITY AND ACHIEVEMENT; DRIVE/ENDURANCE AND MOTIVATION/ENDURANCE; COOPERATION/RESPONSIVENESS/COOPERATION AND RESPONSIVENESS; CONDUCT AND ATTITUDE/GROOMING AND ATTIRE; STAFFING/ORGANIZATION AND STAFFING; Total Number of 3 Weights (Average); REPUTE; PROFESSIONAL AND TECHNICAL SKILLS/TECHNICAL SKILLS/PROFESSIONALISM; PLANNING; CREATIVITY AND INITIATIVE/INITIATIVE; and POTENTIAL.

In summary, it can be concluded that the two short-cut indexing methods, although not achieving the classification accuracy of the original lengthy indexing procedure which had more variables available for the stepwise discriminant analysis process, did, however, achieve an acceptable level of classification performance, that is, approximately comparable to that achieved between Steps 10 and 20 by the longer, more complex indexing methodology. Of the two short-cut methods, the rational condensation indexing method is preferred since it tracked the lengthy method more faithfully in the selection of discriminating variables. Further, the rational condensation method examines *all* of the information contained in a narrative performance evaluation whereas the statistically selected subset method ignores certain portions of the narrative text. Since the rational condensation method is less vulnerable to indexing error and inconsistency and probably will fare better in any cross validation study because it takes into account all of the narrative text, this short-cut method was chosen to be used in further research investigations.

The efficiency of using the rational condensation short-cut indexing method compared to the original lengthy indexing procedure was estimated by considering the time required to index and code both an Evaluation Report with brief narrative comments and one with lengthy narrative comments. The short case contained 43 words and required three minutes to index regardless of which indexing procedure was used. Another half minute was taken up for both procedures in counting the number of words in the text. The process of transferring the indexing decisions and word count to the indexing form and generating the various quantitative variables required approximately two and one-half minutes for each indexing procedure. Entering this information onto IBM coding forms consumed four minutes for the short-cut method (two lines of coding) and six minutes for the lengthy procedure (four lines of coding, the second and fourth lines of which match up with the fields for the first and third lines thus speeding up the coding of the second and fourth lines). In the long case, which contained 338 words, it required 35 minutes to index the narrative text using the lengthy indexing procedure whereas the rational condensation method required only 20 minutes. The additional time required by the lengthy procedure was caused by the difficulty in choosing among the larger number of index terms available for the lengthy method and the need to refer more often to the dictionary of index terms in order to resolve indexing dilemmas. Another five minutes was taken up for both procedures in counting the number of words in the text. The process of transferring the indexing decisions and word count to the indexing form and generating the various quantitative variables required approximately 12 minutes for each indexing procedure. As with the short case, entering the information onto IBM coding forms consumed four minutes for the short-cut method (two lines of coding) and six minutes for the lengthy procedure (four lines of coding, the second and fourth lines of which match up with the fields for the first and third lines thus speeding up the coding of the second and fourth lines).

The conclusion to be drawn from these two comparisons is that there is very little difference between the two indexing procedures in the time required to index and code Evaluation Reports containing brief narrative text. Only when the text becomes longer and requires more scrutiny and consideration by the indexer does the efficiency of the short-cut method become apparent. Over a large sample of Evaluation Reports, it is estimated that use of the rational condensation short-cut indexing method will save 25 to 50 percent of the indexing time required by the original lengthy indexing procedure. It is expected that the time required to count the number of words in the narrative text and to transfer this count and the indexing decisions to the indexing form and to generate the various quantitative variables will be approximately the same for both procedures. The time required for the rational condensation method to enter this information onto IBM coding forms preparatory to keypunching is estimated to be two-thirds of that required by the lengthy indexing procedure. Since only one punched card is needed to contain the variables extracted by the rational condensation content analysis compared to two punched cards for the original lengthy content analysis, the keypunching, verifying, and proofing time is cut in half. And since fewer card images have to be examined by the stepwise discriminant analysis procedure each time that a classification matrix is computed and printed, it is estimated that computer processing time is halved. A more detailed comparison of the efficiency of the two indexing procedures is being carried out as a future area of investigation (see Section 6.D.).

Three samples of Evaluation Reports, covering two contiguous years and representing four occupational specialties and three experimental content analysis procedures, have highlighted certain key variables as being crucial in differentiating between the performance of superlative chief petty officers and their slightly less qualified colleagues. These key variables are the adjectives and adverbs that an evaluator uses to describe the performance of the individual that is being evaluated; the range of skills and abilities that an individual manifests; and the following specific demonstrated capabilities: Management and supervisory ability; skill in leading and directing his men; ability to organize his work area and to staff it properly; ability to plan his workload and take any corrective measures necessary to compensate for unforeseen obstacles to good performance; the ability to present an effective image of his work force to other components of the Navy and to the civilian community; skill in communicating effectively with others; a cooperative and responsive way of performing his job duties; a creative, resourceful, and innovative approach to his work; the drive and stamina to perform well under tiring or adverse circumstances; his level of intellectual functioning; professional and technical competence in his occupational specialty; his level of productivity and achievement; and recognition of his assets and potential by his subordinates, peers, and superior officers.

SECTION 5. EXTENSION OF THE ORIGINAL INTER-INDEXER RELIABILITY STUDY

In the pilot content analytic study of the narrative sections of Navy performance evaluations for senior personnel in Pay Grade E7, the issues of reliability and trainability were of concern although the scope of the small initial research effort did not permit these aspects to be studied in any substantial way. Therefore, in designing the second investigation these issues were dealt with by including a reliability study whose objectives were twofold: (1) to determine the level of agreement among four individuals all of whom independently would perform a content analysis of the same corpus of Evaluation Reports, and (2) to investigate if nonresearchers could be trained successfully to apply the complex content analysis methodology developed in the pilot study.[5] Product-moment correlation, kappa, and weighted kappa were the three statistics used to measure agreement among the four reliability indexers. Of the six possible pairwise comparisons between the four reliability indexers, the value of the various agreement statistics ranged from .64 to .88. The initial expectation in beginning this reliability study was that it would be extremely difficult to train nonresearch-oriented individuals to consistently index the narrative sections of Evaluation Report forms using the complex content analysis methodology that had been developed in the pilot study. The surprising result was that in only six training sessions a quite respectable level of agreement was achieved. This is a significant finding because it suggests that Navy and civilian operational personnel also can be trained to consistently apply content analytic techniques.

In the follow-on investigation to the pilot study and the second study, the original inter-indexer reliability study performed as part of the second study was extended in order to elucidate more fully the question of reliability of the complex, lengthy indexing procedure. The original plan for the extension of the inter-indexer reliability study was to select and train four more individuals in the complex indexing procedure and to have them independently index the same 48 Evaluation Reports that formed the indexing corpus for the first reliability study. However, the results from the first reliability study strongly suggested that additional training of the original reliability indexers aimed at clarifying the areas of confusion that were identified in the analysis of their indexing judgments most likely would raise their level of agreement. Consequently, both of these avenues of investigation were pursued. A revision of the original training manual was prepared by the experienced indexer and the principal investigator, an updated version that attempted to eliminate areas of confusion brought to light in analyzing the results of the first reliability study and which also included voluminous examples of how to handle difficult indexing decisions.[17] This revision was used to train the four participants in the extension of the original reliability study.

Two new reliability indexers were engaged for the study, a male and a female, both in their sophomore year in college. The other two indexers participating in the study were inexperienced indexer A (with two years of college in the liberal arts) and inexperienced indexer B (with executive secretary experience) who also had participated in the first reliability study. All four of these individuals were trained intensively by the experienced indexer over the course of six training sessions using the updated version of the training manual and a corpus of training Evaluation Reports. The two new reliability indexers independently indexed the same 48 Evaluation Reports that were indexed in the first reliability study. These two individuals in essence were attempting to replicate the earlier results. Inexperienced indexer A and inexperienced indexer B were given a new and different set of 48 Evaluation Reports to index independently. This second corpus constitutes a randomized representative sample taken from the cross validation and generalization data bases. This second aspect of the reliability study was included in order to test the hypothesis that with additional training and indexing experience, the level of indexing agreement can be raised.

In each of the two sets of 48 Evaluation Reports, the Evaluation Section was separated from the Justification Section so that the narrative comments for each section of a report were not considered together. This resulted in a group of 96 randomized pieces of narrative text---*minidocuments*---to be indexed by each reliability indexer. To each of these 96 pieces of narrative text was appended the corresponding sections 4A and 4B of the Evaluation Report form. These two sections provide a description of the primary and collateral duties of the individual being evaluated and should be read as background information before beginning to index the narrative text.

When all four reliability indexers had completed indexing their assigned 96 pieces of narrative text, their indexing decisions were recorded on work sheets for each segment of narrative text indexed. These work sheets provided the data base for computing agreement statistics. In all of the statistical computations reported subsequently in this section, assignment of the index terms was considered to be a separate intellectual task from assigning the corresponding weights based on the modifying adjectives and adverbs. There is good justification for analyzing the reliability study results in these two contexts. When an indexer studied a segment of narrative text, the first step was to select an appropriate index term or terms from among the 29 possibilities that best described the substantive content of the text. Once the indexer had completed this first phase of the content analysis, then the segment of narrative text was rescanned to identify the adjectives and adverbs that defined the numerical weight to be assigned to each index term chosen. Considering these judgments as two sequential decision processes also made the results of the reliability study more amenable to statistical analysis as will be shown in the subsequent discussion.

As early as 1960 Cohen, in introducing a new agreement statistic called kappa, pointed out that for most problems in nominal scale agreement between two judges or decision makers, many investigators compute a contingency chi square as a test of the hypothesis of chance agreement, and some investigators have gone on to compute the contingency coefficient, C, as a measure of degree of agreement.[18] However, Cohen concluded that the use of chi square (χ^2), and therefore, the C which is based on it for the evaluation of agreement is indefensible. When applied to a contingency table, χ^2 tests the null hypothesis with regard to *association*, not agreement. Therefore, χ^2 and C are inappropriate statistics for measuring agreement since they will be inflated quite impartially by any departure from chance association, *either disagreement or agreement*. In order to remedy this situation, Cohen suggested a new coefficient, kappa, to measure the degree of agreement in nominal scales, and to provide means for testing hypotheses and setting confidence limits for this coefficient.

Quoting from Cohen's 1960 article [18, pp. 39-40], "...for any problem in nominal scale agreement between two judges, there are only two relevant quantities:

p_o = the proportion of units in which the judges agreed

p_c = the proportion of units for which agreement is expected by chance.

The test of agreement comes then with regard to the $1 - p_c$ of the units for which the hypothesis of no association would predict disagreement between the judges. This term will serve as the denominator.

"To the extent to which nonchance factors are operating in the direction of agreement, p_o will exceed p_c; their difference, $p_o - p_c$, represents the proportion of the cases in which beyond-chance agreement occurred and is the numerator of the coefficient.

"The coefficient κ is simply the proportion of chance-expected disagreements which do not occur, or alternatively, it is the proportion of agreement *after* chance agreement is removed from consideration:

$$\kappa = \frac{p_o - p_c}{1 - p_c} \text{ ."}$$

The significance of an obtained κ is determined by dividing κ by σ_{κ_o} where $\sigma_{\kappa_o} = \sqrt{\frac{p_c}{N(1 - p_c)}}$. The resulting critical ratio is referred to the normal curve. However, Cohen has pointed out that it is generally of as little value to test κ for significance as it is for any other reliability coefficient ---to know merely that κ is beyond chance is trivial since one usually expects much more than this in the way of reliability in psychological measurement. However, the size of the critical ratio does provide some immediate feedback concerning the magnitude of the agreement achieved beyond the level expected by chance. Probably a more useful way to interpret the significance of an obtained κ is in terms of the maximum value of κ. The theoretical upper limit of κ is +1.00, but this limit can only be reached if the off-diagonal (disagreement) cells in the agreement matrix are all zero. This in turn demands that the marginal probabilities for each diagonal (agreement) cell must be identical. Perfect agreement between two judges is rarely achieved, and therefore, the marginal distributions in any agreement matrix are not identical. This means that in practice the upper limit of κ is never +1.00 but rather some lesser value. The maximum value of κ is set by the marginal distributions in any particular application of the kappa agreement statistic, and it can be calculated. A comparison of the obtained κ with its maximum upper limit computed from the marginal distributions provides the investigator with a more useful index of how closely the agreement level that was achieved between two judges approached the maximum level of agreement that was possible.

The kappa statistic was the measure of agreement used in analyzing the index terms assigned by the four reliability indexers and the experienced indexer.

For each segment of narrative text, each indexer chose a term or terms from the list of 29 possibilities, or the decision was made that no term should be used. From a careful analysis of these indexing decisions on the same reliability study data base for each pair of reliability indexers, six pairwise agreement matrices were constructed. These were 30 by 30 matrices, with the 29 index terms representing 29 of the 30 nominal categories and No Index Term Used representing the 30th nominal category. The pairwise indexing decisions for each segment of narrative text analyzed across all 96 documents in the two reliability study data bases were tabulated into the appropriate cell of the agreement matrix for the particular pair of indexers being compared. The 30 diagonal cells of the agreement matrix denote agreement between the two indexers in assigning index terms; all of the off-diagonal elements in the matrix represent instances in which the two indexers disagreed in their selection of terms. The total number of entries in these six matrices varied slightly among the six pairwise comparisons between the five indexers participating in the second reliability study, but in all instances they were very large, ranging from 1,257 tallies to 1,403 tallies. Consequently, the size of the two reliability study data bases can be considered to be large enough to provide a stable measure of the level of agreement achieved in performing this complex intellectual task.

Table 29 shows the results of the kappa analysis of the six pairwise comparisons between the five indexers participating in the second reliability study in selecting index terms for the two reliability study data bases. The second column in this table shows the value of κ; the third column shows the standard error of κ; and the fourth column lists the normal deviate, z, obtained by dividing κ by its standard error. All of the z values are very large, and consequently, extremely significant, indicating that in all six comparisons the null hypothesis that the obtained κ does not exceed the chance level of agreement can be rejected. The fifth column in Table 29 provides the maximum possible value of kappa for each of the six pairwise comparisons. These values can be used as an upper limit for comparing the level of agreement actually achieved with the maximum level possible given the marginal distributions. Thus, in the first comparison for the original data base, that between the experienced indexer and inexperienced indexer X, the κ obtained was .68 compared to a possible maximum value of .88. The last column of Table 29 shows, in percentage form, the ratio of each κ obtained to its maximum value. The best agreement in selecting index terms for the replication of the reliability study on the original data base was obtained between the experienced indexer and inexperienced indexer X (the male college sophomore), a κ of .68. In the previous reliability study all three reliability indexers exceeded this level of agreement with the experienced indexer, demonstrating a range of values for kappa from .72 to .88.

It is rather difficult to speculate just why the two new reliability indexers who were trying to replicate the earlier results did not agree as closely with the experienced indexer on the original reliability data base as the two initial reliability indexers did. Both of the two new indexers were college students, involved in a myriad of academic and recreational activities. It was frustrating to try to schedule the six training sessions because these two individuals had so many conflicts, and sometimes they would cancel at the last minute, making it necessary to reschedule the training session for everyone for another time slot. Because of these considerations, it was felt that perhaps the new indexers were less motivated and not as deeply involved in the second reliability study as the other two indexers who had participated

TABLE 29

RESULTS OF THE KAPPA ANALYSIS FOR THE SIX PAIRWISE COMPARISONS BETWEEN THE FOUR RELIABILITY INDEXERS AND THE EXPERIENCED INDEXER IN SELECTING INDEX TERMS FOR THE TWO RELIABILITY STUDY DATA BASES

Pairwise Comparisons Between Each Pair of Reliability Indexers	κ	σ_{κ_o}	z^*	κ_{max}	κ obtained/ κ_{max} as %
Original Data Base					
The experienced indexer vs. inexperienced indexer X	.68	.0070	98.05	.88	77%
The experienced indexer vs. inexperienced indexer Y	.61	.0066	92.25	.87	70%
Inexperienced indexer X vs. inexperienced indexer Y	.55	.0066	82.81	.82	67%
Second Data Base					
The experienced indexer vs. inexperienced indexer A	.83	.0068	121.78	.95	87%
The experienced indexer vs. inexperienced indexer B	.72	.0065	110.46	.94	77%
Inexperienced indexer A vs. inexperienced indexer B	.70	.0065	106.85	.92	76%

* A z of 3.29 is significant at the .001 level of probability. Therefore, all of the observed values of κ reported in this table are extremely significant and lead to rejection of the null hypothesis that the obtained κ does not exceed the chance level of agreement.

in the first study conducted a year earlier. These latter two individuals are regular employees of R-K Research and System Design, performing a variety of clerical and technical assignments in addition to their role in the two reliability studies. This explanation may account for the observed differences in the level of agreement with the experienced indexer by indexers X and Y in the second study and indexers A and B in the first study.

Turning attention now to the second part of Table 29---the results of the kappa analysis for the second reliability study data base, the best agreement in selecting index terms was obtained between the experienced indexer and inexperienced indexer A, a κ of .83. In the first reliability study this same indexer also demonstrated the highest level of agreement with the experienced indexer in selecting index terms, a κ of .88. The level of agreement between the experienced indexer and inexperienced indexer B in assigning index terms as measured by κ was .72 in both of the reliability studies. Thus, neither inexperienced indexer A nor inexperienced indexer B was able to increase her level of agreement with the experienced indexer despite refresher training in the complex, lengthy indexing procedure and the challenge to try to outdo her previous performance. However, these two reliability indexers felt that the data base indexed by them in the second reliability study contained a sample of narrative text more difficult to index than the first reliability study data base, and the experienced indexer who conducted the refresher training concurred in this judgment. There were more longer cases in the second reliability study data base, and in addition, the wording of the narrative text in general was more complex, confusing, or vague. Both reliability indexers expressed difficulty in trying to understand what the evaluators meant, and even after several reviews of their indexing decisions over the entire second reliability study data base, they remained uncertain of their choice of many index terms. As a result it also took longer for them to index the second reliability study data base. Therefore, the greater difficulty in indexing this narrative material may have masked any gain in indexing proficiency that might have been achieved by the additional training. Another possible explanation is that inexperienced indexers A and B may have already approached the upper boundary of their indexing skill, with additional training and experience contributing very little to increasing their level of agreement with the experienced indexer. Regardless of which explanation one accepts as being more plausible to account for the results, the reassuring finding is that in only six training sessions, once again a fairly respectable level of agreement among indexers was achieved for both reliability study data bases.

The kappa analysis performed as part of the first reliability study revealed that the major area of confusion in indexing the initial reliability study data base resided in whether or not to index supposedly factual statements describing the job duties and the qualifications needed for the position occupied by the person being evaluated rather than this individual's actual performance in the position. All three of the less experienced indexers tended to index these statements as describing the individual's performance whereas the experienced indexer whom the other three indexers were trying to emulate treated these statements as factual descriptions of the job duties and the qualifications needed for the position. It was concluded that additional training aimed at clarifying this area of confusion most likely would markedly reduce this type of disagreement and possibly raise the magnitude of kappa. In order to test this assumption, the six pairwise comparisons shown in Table 29 were recomputed by removing the "No Index Term Used" nominal category from the analysis. Four of the six kappa values remained unaltered by this recalculation, only two of them being affected and both of these being increased by only .01. Therefore, it appears that this area of confusion indeed was resolved in the training sessions for the second reliability study and did not constitute a significant factor in reducing the level of agreement achievable. It also is interesting to note that for both data bases shown in Table 29, the lowest level of agreement was between the two inexperienced indexers, each pair of inexperienced indexers agreeing more closely with the indexing decisions of the experienced indexer than with each other. This is not a surprising finding since the inexperienced indexers were trying to emulate the indexing skill of the experienced indexer who served as the model.

Analysis of the level of agreement among the five indexers participating in the second reliability study in assigning numerical weights to each index term selected, based on the modifying adjectives and adverbs, was performed differently than the analysis of the level of agreement in selecting the index terms themselves. Selection of the index terms in the two reliability studies constituted a nominal scale whereas assignment of a numerical weight to each index term selected was an indexing decision involving an ordinal scale. Therefore, more powerful agreement statistics could be employed. Since numerical weights on a scale from 1 to 5 (New -2 to New 3) were assigned to each index term selected, it was possible to compute a product-moment correlation coefficient between each pair of reliability indexers. The new transformed weights were used in these computations since this ordinal scale provided a more justifiable way of measuring the situation in which one indexer did not select an index term but the other indexer did (see Table 9).

In addition to computing these six product-moment correlation coefficients, another agreement statistic, weighted kappa, also was calculated in order to determine if it agreed with the results of the correlational analysis. In 1968 Cohen published another article generalizing the kappa statistic to the situation in which disagreements of varying gravity can be weighted accordingly.[19] Application of weighted kappa to quantifying the level of agreement in psychiatric diagnosis also was shown by Cohen and his colleagues.[20]

Weighted kappa is an agreement statistic corrected for chance agreement, to be used when different kinds of disagreement are to be differentially weighted in the agreement index. The desired weighting is accomplished by an a priori assignment of weights to the *r* by *c* cells of the agreement matrix, and must be done very carefully because the weights assigned are an integral part of how agreement is defined, and therefore, how it is measured with weighted kappa (κ_ω). Table 30 shows the weighting algorithm that was used in computing κ_ω for assessing the level of agreement in assigning numerical weights, based on the modifying adjectives and adverbs, to the index terms selected in the two reliability studies. The first step in computing κ_ω was to construct a 6 by 6 agreement matrix between each pair of reliability indexers that encompassed all of the pairwise numerical weights that were assigned to index terms based on their modifying adjectives and adverbs. These numerical weights were tabulated in the agreement matrix across all 96 documents in each reliability study data base. Using the first row of Table 30 as an example, if reliability indexer I and reliability indexer II both had assigned a numerical weight of 3 to the index term that they had selected, it represented perfect agreement in their interpretation of the superlativeness of the adjective or adverb modifying the index term. Therefore, the 3,3 cell was given an a priori weight of zero in computing κ_ω since perfect agreement should receive no penalty. If one indexer had assigned a numerical weight of 3 to the index term selected and the other indexer had assigned a numerical weight of 2, they only disagreed by one position on the ordinal scale, and therefore, the 3,2 cells were given an a priori weight of one in computing κ_ω, penalizing this mild disagreement only slightly. In the extreme case, if one indexer had assigned a numerical weight of 3 to the index term selected and the other indexer had assigned a numerical weight of -2, they disagreed by five positions on the ordinal scale, and therefore, the 3,-2 cells were given an a priori weight of five in computing κ_ω, penalizing this extreme disagreement the maximum possible. This same logic was applied in determining the weights to be used in computing κ_ω through-

TABLE 30

THE WEIGHTING ALGORITHM USED IN COMPUTING WEIGHTED KAPPA FOR ASSESSING THE LEVEL OF AGREEMENT IN ASSIGNING NUMERICAL WEIGHTS TO THE INDEX TERMS SELECTED IN THE TWO RELIABILITY STUDIES

	RELIABILITY INDEXER I Index Weights					
RELIABILITY INDEXER II Index Weights	3	2	1	0	-1	-2
3	*0*	*1*	*2*	*3*	*4*	5
2	*1*	*0*	*1*	*2*	*3*	*4*
1	*2*	*1*	*0*	*1*	*2*	*3*
0	*3*	*2*	*1*	*0*	*1*	*2*
-1	*4*	*3*	*2*	*1*	*0*	*1*
-2	5	*4*	*3*	*2*	*1*	*0*

out the remainder of the matrix. All of the diagonal cells were given a weight of zero since in no case should perfect agreement be penalized. All cells immediately off the diagonal were penalized by a weight of one; those cells slightly farther off the diagonal were penalized by a weight of two; and so on out to a penalty weight of five for the case of worst disagreement.

The formula for computing κ_ω is

$$\kappa_\omega = 1 - \frac{\Sigma w_{ij} p_{oij}}{\Sigma w_{ij} p_{cij}}$$

where w_{ij} = a priori weight in cell *ij*

p_{oij} = observed proportion in cell *ij*

p_{cij} = chance proportion in cell *ij*.

The original approximate standard error formula* for κ_ω is

$$\sigma_{\kappa_{\omega o}} = \sqrt{\frac{\Sigma w_{ij}{}^2 p_{cij} - (\Sigma w_{ij} p_{cij})^2}{N(\Sigma w_{ij} p_{cij})^2}} .$$

A significance test of κ_ω, that is, a test of H_o: Population κ_ω - Observed $\kappa_\omega = 0$, is accomplished by evaluating the normal curve deviate

$$z = \frac{\kappa_\omega}{\sigma_{\kappa_{\omega o}}} .$$

It is possible to capitalize on the fact that the standard normal deviate squared is distributed as χ^2 with one *df* and then to cast the result in the χ^2 form traditionally used in the analysis of frequencies and proportions:

$$z^2 = \chi^2_\omega = \left(\frac{\kappa_\omega}{\sigma_{\kappa_{\omega o}}}\right)^2 ,$$

where χ^2 is distributed as χ^2 with *one* degree of freedom, no matter what the dimensionality of the agreement matrix is. The formula for χ^2_ω (weighted chi square) is offered by Cohen for use in all contexts where χ^2 is now used with frequency and proportion data, and where the investigator wishes to improve the power of the statistical test by including his hypotheses (hunches, expectations) about the outcome. Weighted kappa is an incidental benefit in this scheme in that it provides a measure of hypothesized association, a "rho" measure.[23]

The more recent publications[21,22] on kappa and weighted kappa provide computational examples that use an *agreement* weighting scale, and the corresponding calculation of the standard error reflects this way of scaling the analysis. However, either degree of agreement or degree of disagreement may be scaled, depending on what weighting scheme seems more natural in a given context. The weighting scheme for κ_ω in the original reliability study and in the extension of this study was developed in terms of *disagreement* scaling. Therefore, to switch now to another scaling logic based on agreement rather than disagreement would make it impossible to directly compare the results of the second reliability study with those of the first study. Consequently, it was decided to retain the disagreement weighting algorithm shown in Table 30. Since all of the observed weighted kappas in both of the reliability studies were very large, and as a result, significantly different from zero far beyond the .001 level of probability, the fact that the conservative formula for estimating the standard error of κ_ω overestimated these values really has no effect on the interpretation of the statistical results. The same conclusions would be arrived at regardless of which formula for estimating the standard error was used. For example, there is no doubt that a weighted kappa of .63, having a z value of 119.62 and a corresponding chi square value of 14,308.2 with *one* degree of freedom is extremely significant when the values of z and χ^2_1 at the .001 level of probability are 3.29 and 10.8, respectively. As pointed out by Cohen, a more meaningful way to interpret an observed value of kappa is to compare it to its maximum upper limit in order to assess how closely the agreement level that was achieved between two judges actually approached the maximum level of agreement that was possible.[18] However, the weights which maximize *weighted* kappa turn out to be of no psychological interest because κ_ω is a maximum only for binary weights assigned so that the cell where $p_{ij}/p_{i.}p_{.j}$ is smallest in the agreement matrix is assigned a weight of 0 and *all* the other cells in the matrix are assigned a weight of 1.[22]

---

* In two more recent publications[21,22], Cohen indicated that the original approximate standard error formula for κ and κ_ω published in his earlier articles[18,19] were incorrect, but in a conservative direction, i.e., too large.

Since the correct version of the formula for the standard error of κ_ω is so cumbersome to compute and would have only *increased* the magnitude of the resulting z values and corresponding χ^2_1 values which are enormous anyway as computed by the less accurate, more conservative formula for the standard error, it was felt that the extensive additional computation required to redo the entire κ_ω analysis for both reliability studies based on agreement scaling rather than on the disagreement weighting scheme developed in the first reliability study was not warranted, the interpretation of the results being the same regardless of which formula was used to compute σ_{κ_ω}. With these statistical considerations kept in mind, Table 31 now can be discussed.

Table 31 shows the results of the correlational analysis and the weighted kappa analysis for the six pairwise comparisons between the five indexers participating in the second reliability study in assigning numerical weights to each index term selected, based on the modifying adjectives and adverbs, for the two reliability study data bases. The results of the correlational analysis are shown first in Table 31. The best agreement in assigning numerical weights to each index term selected for the replication of the reliability study on the original data base was obtained between the experienced indexer and inexperienced indexer X, a correlation coefficient of .60. The best agreement in selecting index terms themselves also was achieved between this same pair of indexers (see Table 29). In the previous reliability study all three reliability indexers exceeded this level of agreement with the experienced indexer, demonstrating a range of correlation coefficients from .64 to .80. A difference in their involvement in the replication of the first reliability study by the two new indexers compared to the two reliability indexers who participated in the first study appears to be the best explanation to account for these findings.

On the weighted kappa side of Table 31 for the original data base, the best agreement in assigning numerical weights to each index term selected as measured by weighted kappa again was obtained between the experienced indexer and inexperienced indexer X, a κ_ω of .65. The values shown in parentheses after the four κ_ω's listed in Table 31 were computed in order to determine the level of agreement achieved if those instances were excluded where the experienced indexer did not select an index term, and consequently, did not assign a

TABLE 31

RESULTS OF THE CORRELATIONAL ANALYSIS AND THE WEIGHTED KAPPA ANALYSIS FOR THE SIX PAIRWISE COMPARISONS BETWEEN THE FOUR RELIABILITY INDEXERS AND THE EXPERIENCED INDEXER IN ASSIGNING NUMERICAL WEIGHTS TO EACH INDEX TERM SELECTED FOR THE TWO RELIABILITY STUDY DATA BASES

Pairwise Comparisons Between Each Pair of Reliability Indexers	Product-Moment Correlation r_{pm}*	Weighted Kappa			
		κ_ω**	$\sigma_{\kappa_{\omega o}}$	z***	χ^2_1***
Original Data Base					
The experienced indexer vs. inexperienced indexer X	.60	.65 (.69)	.0101	64.57	4,169.4
The experienced indexer vs. inexperienced indexer Y	.54	.55 (.63)	.0122	44.95	2,020.1
Inexperienced indexer X vs. inexperienced indexer Y	.50	.48	.0143	33.82	1,143.5
Second Data Base					
The experienced indexer vs. inexperienced indexer A	.74	.76 (.78)	.0081	93.66	8,772.1
The experienced indexer vs. inexperienced indexer B	.62	.63 (.70)	.0053	119.62	14,308.2
Inexperienced indexer A vs. inexperienced indexer B	.58	.59	.0054	109.43	11,975.3

* A product-moment correlation coefficient of .104, based on an N of 1000, is significantly different from zero at the .001 level of probability.

** The values of κ_ω shown in parentheses take into account only those instances in which both indexers selected an index term and exclude those instances in which the experienced indexer did not select an index term but the other less experienced indexer did.

*** A z of 3.29 and a χ^2_1 of 10.8 are significant at the .001 level of probability. Therefore, all of the observed values of κ_ω reported in this table are highly significant and lead to rejection of the null hypothesis that the observed κ_ω does not exceed the chance level of agreement.

numerical weight but the other less experienced indexer did select an index term and assigned a weight to it. This had proved to be the area of major confusion in executing the first reliability study as pointed out earlier in this section in discussing the results of the kappa analysis of level of agreement in selecting the index terms themselves. Instances where the experienced indexer did not assign a weight but the other indexer did form one row in the weighted kappa computational matrix. This row can be omitted from the computation, resulting in a value for κ_ω that ignores this area of confusion and takes into account only those instances where both indexers selected an index term, and consequently, assigned a weight. The gain in the value of κ_ω is not very large for the comparison between the experienced indexer and inexperienced indexer X when κ_ω was recomputed in this fashion. However, the gain was considerable in the comparison between the experienced indexer and inexperienced indexer Y, indicating that inexperienced indexer Y was more prone to overindex than inexperienced indexer X. As was expected, the κ_ω values in Table 31 are similar in magnitude to their correlation coefficient counterparts. All measures of agreement shown in Table 31 are significantly different from zero well beyond the .001 level of probability.

Turning attention now to the second part of Table 31---the results of the correlational and weighted kappa analysis for the second reliability study data base, the best agreement in assigning weights to the index terms selected again was obtained between the experienced indexer and inexperienced indexer A, a correlation coefficient of .74 with a corresponding weighted kappa of .76. In the first reliability study this same indexer also demonstrated the highest level of agreement with the experienced indexer in assigning weights to the index terms selected, a correlation coefficient of .80 with a corresponding weighted kappa of .78. The level of agreement between the experienced indexer and inexperienced indexer B in assigning weights to the index terms selected was approximately the same in both studies, a correlation coefficient of .64 with a corresponding weighted kappa of .60 in the first reliability study and a correlation coefficient of .62 with a corresponding weighted kappa of .63 in the second reliability study which involved a different data base. Thus, neither inexperienced indexer A nor inexperienced indexer B was able to increase her level of agreement with the experienced indexer in assigning weights despite additional training and a high motivational level. The explanation for this outcome appears to be either the greater difficulty of the second data base or that these two individuals have asymptotically approached their best performance in a content analysis task as complex as this one. The recomputation of κ_ω to exclude those instances in which the experienced indexer did not select an index term but the other less experienced indexer did does not raise the magnitude of weighted kappa very much for inexperienced indexer A but does show some gain for inexperienced indexer B, indicating that inexperienced indexer B was more prone to overindex than inexperienced indexer A. A possible reason that inexperienced indexer A exhibited the best agreement with the experienced indexer in both of the reliability studies is that in other aspects of this research she was responsible for entering the indexing decisions of the experienced indexer for the pilot study, cross validation, and generalization samples onto indexing forms and then onto IBM coding forms preparatory to keypunching. Although the rationale for the content analysis methodology was not explained to her at that time because she did not need to understand it in order to perform what was essentially a clerical coding operation, it is quite possible that she absorbed the logic of the indexing scheme by example and that her coding duties served the purpose of providing her with an extended additional training period in the content analysis methodology.

In summary, the conclusions that can be drawn from this extension of the original reliability study are that once again, in only six training sessions, a fairly respectable level of agreement was achieved on a very difficult content analysis task. The two new reliability indexers (both college sophomores) who were attempting to replicate the results from the first study did not achieve as high a level of agreement with the experienced indexer as the three reliability indexers did in the initial study, probably because the two new indexers were less motivated and not so deeply involved in the second reliability study as inexperienced indexers A and B were in the first study conducted a year earlier. These latter two individuals are regular employees of R-K Research and System Design, performing a variety of clerical and technical assignments in addition to their role in the two reliability studies. Inexperienced indexer A in particular may have had additional unsuspected training in the content analysis methodology since one of her other assignments in this research was to enter the indexing decisions of the experienced indexer for the pilot study, cross validation, and generalization samples onto IBM coding forms for keypunching. Inexperienced indexer A's extended exposure to the logic of the indexing scheme in the context of preparing the coding forms may account for her superior performance in both reliability studies.

In that part of the second reliability study designed to test the hypothesis that with additional training and indexing experience the level of indexing agreement can be raised, the results were ambiguous. Neither inexperienced indexer A nor inexperienced indexer B was able to increase her level of agreement with the experienced indexer despite refresher training in the complex, lengthy indexing procedure and the challenge to try to outdo her previous performance. However, these two reliability indexers felt that the data base indexed by them in the second reliability study contained a sample of narrative text more difficult to index than the first reliability study data base, and this greater difficulty inherent in the narrative text may have masked any gain in indexing proficiency that might have been achieved by the additional training. Another possible explanation is that inexperienced indexers A and B may have already approached the upper boundary of their indexing skill, with additional training and experience contributing very little to increasing their level of agreement with the experienced indexer.

SECTION 6. FUTURE AREAS OF INVESTIGATION

The obvious next step in this research is to cross validate the superior short-cut indexing technique---the rational condensation method---on other occupational specialties and on other pay grades than those studied to date (viz., AT's, BT's, CS's, and RM's in Pay Grade E7). In the past year a new performance evaluation report form, NAVPERS 1616/18, for Pay Grades E5 and E6 was introduced into operational use. It will take a year before the raw marks given on this form from an actual operational evaluation can be converted to T scores, a necessary requirement in order to provide valid criterion data for research studies. However, a set of usable fleet trial data exists at NPRDC that was generated in the process of testing a number of experimental forms for measuring on-job performance for Pay Grades E5 and E6.[3] One of these forms, the form recommended by NPRDC, is very similar to NAVPERS 1616/18, the narrative evaluation and justification comments sections being essentially the same. Further, this extensive data base of fleet trial data, which includes useful criterion data, consists of evaluations on enlisted personnel in seven occupational areas, only one of which (Radioman) overlaps the four occupational specialties already studied:

AD - Aviation Machinists Mate
DC - Damage Controlman
ET - Electronics Technician
HM - Hospital Corpsman
PN - Personnelman
RM - Radioman
SK - Storekeeper

Until the statistical standardization of the E5-E6 evaluations on NAVPERS 1616/18 become available, the fleet trial data described above offer an immediate opportunity to cross validate the superior short-cut content analysis technique that has been developed.

During the 12-month time period from January 1, 1974 to December 31, 1974, additional validity and reliability studies of content analysis techniques for extracting differentiating information from narrative performance evaluations will be carried out with the ultimate objective that of recommending to NPRDC a method that can be tested in a future convocation of a simulated or actual selection board. The following specific tasks are being undertaken:

A. Cross Validation of the Rational Condensation Short-cut Indexing Procedure on the E5-E6 Fleet Trial Data

A sample has been selected that is representative of the seven occupational specialties (enumerated above) from the E5-E6 fleet trial data base at NPRDC. The narrative performance evaluation and justification comments contained in this sample are being indexed using the rational condensation short-cut indexing procedure. Stepwise discriminant analysis will be used to determine how well the quantitative variables derived from the short-cut content analysis of the narrative text can classify each individual evaluated into correct criterion group. Each of the seven occupational specialties represented in the sample and each of the two pay grades will be analyzed separately. This study will show if the rational condensation short-cut indexing procedure is generalizable to Pay Grades E5 and E6 and to occupational specialties other than those studied thus far.

B. Reliability Study of the Rational Condensation Short-cut Indexing Procedure

A third reliability study is being conducted in order to be certain that consistency among several indexers can be taught and achieved in their interpretation and application of the rational condensation short-cut indexing procedure. A new training manual is being prepared to explain and illustrate the proper utilization of this short-cut technique. This manual will be used to train three reliability indexers. When their training is completed, they independently will index the narrative comments contained in a newly selected set of 48 Evaluation Reports from the E5-E6 fleet trial data base. The level of agreement between each of the three reliability indexers and the experienced indexer who trains them will be determined by the same statistical procedures used in the two earlier reliability studies in order that comparisons can be made among the three reliability studies of the magnitude of agreement that was achieved. This study will lay the foundation for a training curriculum that may be used in the future to train Navy and civilian operational personnel in the application of the content analysis methodology.

C. Validation of the Original Indexing Procedure by Means of a Second Indexer

The results of the first reliability study suggested the possibility that it may be as important to consider the issue of internal consistency for a single indexer as to measure the level of agreement that can be achieved among several indexers. It seems reasonable to assume that although there may be slight differences between two indexers in how they apply a particular indexing procedure, a more important consideration is that they consistently use their own individualized interpretation of the indexing rules and conventions. One then might expect that regardless of which individualized interpretation was used to index a particular data base, a similar level of classification agreement with the criterion of on-job performance could be achieved. This is an important area to study because the findings may point to the necessity to use only one indexer for a particular data base if optimum extraction of differentiating information is to be obtained.

In order to shed some light on this issue, a second indexer is independently reindexing the cross validation and generalization samples. Thus, an exact replication of the indexing performed by the experienced indexer in her content analysis of the cross validation and generalization samples is being carried out independently. The accuracy of classification into correct criterion group achieved by each of these two indexers will be compared in order to determine if both indexers working separately with their own individualized interpretations of the indexing rules and conventions can achieve comparable classification results. The original lengthy indexing procedure is being used in this comparison because the experienced indexer's indexing decisions and the classification results based on her judgments are immediately available for this study.

D. Efficiency of the Rational Condensation Short-cut Indexing Procedure Compared to the Original Lengthy Indexing Procedure

A careful comparison of the indexing, coding, keypunching, and subsequent computer processing time required to apply the rational condensation short-cut indexing procedure and the original lengthy indexing procedure to a small subsample of the E5-E6 fleet trial data base will be made. This comparison will provide data needed for assessing the economic feasibility of adding information extracted from narrative comments into a composite score for predicting an enlisted man's potential for assuming the managerial responsibilities of the next higher pay grade.

REFERENCES

1. Robertson, D. W. Source documents for the automated enlisted performance evaluation system. Research Memorandum SRM 72-10, March 1972, Naval Personnel and Training Research Laboratory, San Diego, California 92152.

2. Robertson, D. W., James, J., & Royle, Marjorie H. Feasibility of computer-generated data displays in the automated performance evaluation system. Research Report SRR 72-20, April 1972, Naval Personnel and Training Research Laboratory, San Diego, California 92152.

3. Robertson, D. W., Royle, Marjorie H., & James, J. Design and fleet trial of automated performance evaluation forms for two pay grade groups: E5-E6 and E1-E4. Research Report SRR 72-11, November 1972, Naval Personnel and Training Research Laboratory, San Diego, California 92152.

4. Ramsey-Klee, Diane M. A content analysis of the narrative sections of Navy performance evaluations for senior enlisted personnel. Technical Report No. 2-71, October 1971, R-K Research and System Design, Contract Nonr N00014-70-C-0291 - Project NR 153-316, Office of Naval Research.

5. Ramsey-Klee, Diane M., & Richman, Vivian. Cross validation and generalization of a content analysis of the narrative sections of Navy performance evaluations for senior enlisted personnel. Technical Report No. 1-73, April 1973, R-K Research and System Design, Contract Nonr N00014-72-C-0231 - Project NR 150-344, Office of Naval Research.

6. Baxendale, Phyllis. Content analysis, specification, and control. In C. A. Cuadra (Ed.), *Annual Review of Information Science and Technology, Vol. 1.* New York: Interscience Publishers, 1966. Pp. 71-72.

7. Taulbee, O. E. Content analysis, specification, and control. In C. A. Cuadra (Ed.), *Annual Review of Information Science and Technology, Vol. 3.* Chicago: Encyclopaedia Britannica, Inc., 1968. P. 106.

8. Fairthorne, R. A. Content analysis, specification, and control. In C. A. Cuadra (Ed.), *Annual Review of Information Science and Technology, Vol. 4.* Chicago: Encyclopaedia Britannica, Inc., 1969. P. 76.

9. Koontz, H., & O'Donnell, C. *Principles of Management.* (5th Ed.) New York: McGraw-Hill Book Co., 1972.

10. Koontz, H. *Appraising Managers As Managers.* New York: McGraw-Hill Book Co., 1971.

11. Newman, W. H., Summer, C. E., Jr., & Warren, E. K. *The Process of Management.* (2nd Ed.) Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967.

12. Dixon, W. J. (Ed.) *BMD: Biomedical Computer Programs.* Los Angeles: University of California Press, 1973.

13. Ramsey-Klee, Diane M., & Richman, Vivian. Cross validation and generalization of a content analysis of the narrative sections of Navy performance evaluations for senior enlisted personnel. Technical Report No. 1-73, April 1973, R-K Research and System Design, Contract Nonr N00014-72-C-0231 - Project NR 150-344, Office of Naval Research, pp. 42-61.

14. Darlington, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1968, *69*(3), 161-182.

15. Ramsey-Klee, Diane M., & Richman, Vivian. Cross validation and generalization of a content analysis of the narrative sections of Navy performance evaluations for senior enlisted personnel. Technical Report No. 1-73, April 1973, R-K Research and System Design, Contract Nonr N00014-72-C-0231 - Project NR 150-344, Office of Naval Research, p. 73.

16. *Ibid*, p. 79.

17. *Ibid*, pp. 95-150.

18. Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, *20*, 37-46.

19. Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, October 1968, *70*(4), 213-220.

20. Spitzer, R. L., Cohen, J., Fleiss, J. L., & Endicott, Jean. Quantification of agreement in psychiatric diagnosis: A new approach. *Archives of General Psychiatry*, July 1967, *17*, 83-87.

21. Fleiss, J. L., Cohen, J., & Everitt, B. S. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 1969, *72*(5), 323-327.

22. Cohen, J. Weighted chi square: An extension of the kappa method. *Educational and Psychological Measurement*, 1972, *32*, 61-74.

23. Cohen, J. Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of Clinical Psychology*. New York: McGraw-Hill, 1965. Pp. 101-106.